3D Epigenomic Characterization Reveals Insights Into Gene Regulation and Lineage Specification During Corticogenesis

Authors: Michael Song[1,2,*], Mark-Phillip Pebworth[3,4,*], Xiaoyu Yang[1,*], Armen Abnousi[5], Changxu Fan[6,7], Jia Wen[8], Jonathan D. Rosen[9], Mayank NK Choudhary[6,7], Xiekui Cui[1], Ian R. Jones[1], Seth Bergenholtz[1], Ugomma C. Eze[4,9], Ivan Juric[5], Bingkun Li[1], Lenka Maliskova[1], Weifang Liu[9], Alex A. Pollen[4,11], Yun Li[8,9,12], Ting Wang[6,7], Ming Hu[5,*,#], Arnold R. Kriegstein[4,11,#], Yin Shen[1,11,#]

Affiliations

[1]Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA.
[2]Pharmaceutical Sciences and Pharmacogenomics Graduate Program, University of California, San Francisco, San Francisco, CA, USA.
[3]Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA, USA.
[4]The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, UCSF, San Francisco, CA, USA.
[5]Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH, USA.
[6]Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA.
[7]The Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO, USA.
[8]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA.
[9]Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA.
[10]Medical Scientist Training Program, University of California, San Francisco, San Francisco, CA, USA.
[11]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA
[12]Department of Computer Science, University of North Carolina, Chapel Hill, NC, USA.

*These authors contributed equally to the work.
#Corresponding authors: Ming Hu hum@ccf.org, Arnold R. Kriegstein Arnold.Kriegstein@ucsf.edu, and Yin Shen Yin.Shen@ucsf.edu.

47 **Abstract**

48

49 Lineage-specific epigenomic changes during human corticogenesis have previously

50 remained elusive due to challenges with tissue heterogeneity and sample availability.

51 Here, we analyze cis-regulatory chromatin interactions, open chromatin regions, and

52 transcriptomes for radial glia, intermediate progenitor cells, excitatory neurons, and

53 interneurons isolated from mid-gestational human brain samples. We show that

54 chromatin looping underlies transcriptional regulation for lineage-specific genes, with

55 transcription factor motifs, families of transposable elements, and disease-associated

56 variants enriched at distal interacting regions in a cell type-specific manner. A subset of

57 promoters exhibit unusually high degrees of chromatin interactivity, which we term super

58 interactive promoters. Super interactive promoters are enriched for critical lineage-

59 specific genes, suggesting that interactions at these loci contribute to the fine-tuning of

60 cell type-specific transcription. Finally, we present CRISPRview, a novel approach for

61 validating distal interacting regions in primary cells. Our study presents the first

62 characterization of cell type-specific 3D epigenomic landscapes during human

63 corticogenesis, advancing our understanding of gene regulation and lineage specification

64 during human brain development.

65

66 **Introduction**

67

68 The human cortex is a complex, heterogeneous structure that undergoes extensive

69 expansion during development, a process which is markedly different and features

70 distinct cell types from mouse cortical development. Previous studies utilized single cell

71 RNA sequencing (scRNA-seq) to unravel the transcriptomic diversity of the developing

72 cortex, revealing at least nine major cell types and up to 26 distinct subtypes in the dorsal

73 cortex alone[1,2]. Much of this diversity arises from cortical stem cells known as radial glia

74 (RG), whose cell bodies reside in the germinal zones (GZs) of the dorsal and ventral

75 cortex. In the dorsal cortex, RG divide asymmetrically to give rise to intermediate

76 progenitor cells (IPCs), which proliferate and differentiate into excitatory neurons (eNs)[3,4].

77 These newborn neurons undergo radial migration until they reach the cortical plate (CP),

78   where they mature and undergo synaptogenesis[5]. Meanwhile, interneurons (iNs)
79   produced in the ventral cortex migrate tangentially into the dorsal cortex through the
80   marginal and germinal zones[6]. These processes result in a CP consisting primarily of eNs
81   and iNs and a GZ where all four cell types are intermixed.

82

83   Dynamic changes in the epigenomic landscape have been shown to play a critical role in
84   development and cell fate commitment, for instance through the rewiring of physical
85   chromatin loops between promoters and distal regulatory elements[7]. These regulatory
86   interactions are of particular interest as their dysregulation has been linked to complex
87   diseases and traits[8,9]. Despite their utility, detailed epigenomic characterizations are still
88   absent for specific cell types in the developing cortex due to shortcomings associated
89   with the analysis of bulk tissues[10,11]. Here, we present a novel strategy for isolating RG,
90   IPCs, eNs, and iNs from mid-gestational human brain samples, enabling the cell type-
91   specific profiling of their epigenomic features. In addition, we present CRISPRview, a
92   technique for validating distal regulatory regions in primary cells, demonstrating that
93   *GPX3*, *TNC*, and *HES1* are regulated by distal enhancers in RG. Our results identify novel
94   mechanisms underlying gene regulation and lineage specification during corticogenesis,
95   providing a framework for the elucidation of diverse processes in development and
96   disease.

97

98   **Results**

99

100   **Isolation of specific cell populations from the developing human cortex.**

101

102   To isolate four specific cell populations (RG, IPCs, eNs, and iNs) from mid-gestational
103   human brain samples between gestational weeks (GW) 15 to 22 (**Supplementary Table
104   1**), we adapted a previously reported approach for isolating RG from human cortical
105   samples using fluorescence-activated cell sorting (FACS)[12]. Specifically, we incorporated
106   markers for additional cell types from a recently published scRNA-seq dataset in the
107   human neocortex[1]. Microdissected GZ and CP samples were dissociated, stained using
108   antibodies for EOMES, SOX2, PAX6, and SATB2, then partitioned into their constituent

109 populations using FACS (**Fig. 1a and Supplementary Fig. 1**). IPCs were first isolated as

110 the EOMES+ population. eNs were isolated from the EOMES- and SOX2- population

111 based on high expression of SATB2, which marks both newborn and mature eNs at the

112 ages of the samples[1]. RG were isolated based on the high expression of both SOX2 and

113 PAX6, and iNs were isolated based on medium SOX2 and low PAX6 expression.

114

115 The gene expression profiles of the sorted cell populations were both consistent with

116 cellular identity and reproducible between individuals (**Fig. 1b and Supplementary Fig.**

117 **2a**). Sorted RG expressed *VIM*, *HES1*, *GPX3*, and *GFAP*, with little to no expression of

118 marker genes for other cell types, whereas sorted IPCs expressed the IPC marker genes

119 *EOMES*, *SSTR2*, and *NEUROD4*. In concordance with previous reports, *PAX6* was

120 expressed in both RG and IPCs. Sorted eNs expressed the eN marker genes *SLA*,

121 *SLC17A7*, *SATB2*, and *TBR1*, whereas *DLX1*, *DLX2*, and *GAD1* were exclusively

122 detected in sorted iNs. When compared with aggregated scRNA-seq gene expression

123 profiles[1], our sorted cell populations exhibited the highest correlation with their

124 corresponding subtypes while also showing reduced correlation with cells from the

125 endothelial, mural, microglial, and choroid plexus lineages (**Fig. 1c**). Based on these

126 results, we determined that our sorting strategy was robust and our sorted cell populations

127 were suitable for additional epigenomic profiling.

128

129 **Characterization of 3D epigenomic landscapes during corticogenesis.**

130

131 We performed H3K4me3-centric proximity ligation-assisted ChIP-seq (PLAC-seq) to

132 identify chromatin interactions at active promoters, assay for transposase-accessible

133 chromatin using sequencing (ATAC-seq) to demarcate open chromatin regions, and RNA

134 sequencing (RNA-seq) to profile transcriptomes in the sorted RG, IPCs, eNs, and iNs

135 (**Fig. 1d and Supplementary Table 2**). We first confirmed the reproducibility of all PLAC-

136 seq and ATAC-seq replicates (**Supplementary Fig. 2b and 2c**). Next, we applied the

137 MAPS pipeline[13] to call significant H3K4me3-mediated cis-regulatory chromatin

138 interactions in merged replicates for each cell type at a resolution of 5 kb. We identified

139 35,552, 26,138, 29,104 and 22,598 MAPS interactions for RG, IPCs, eNs, and iNs,

140 respectively, with approximately 85% of the interactions classified as anchor to non-
141 anchor (XOR), and the remaining interactions classified as anchor to anchor (AND) (**Fig.**
142 **1e and Supplementary Fig. 3a and 3b**). The median interaction distance was between
143 170 kb to 230 kb for each cell type (**Fig. 1f**), and the majority of interactions occurred
144 within TADs in GZ and CP tissues[10] (**Supplementary Fig. 3c**).

145

146 **H3K4me3-mediated chromatin interactions contribute to cell type-specific gene**
147 **expression.**

148

149 Since H3K4me3 is a histone mark associated with active promoters, we were interested
150 to determine to what extent 3K4me3-mediated chromatin interactions influenced gene
151 expression. First, we observed that the sorted cell populations clustered by
152 developmental age based on their interaction strengths (**Fig. 2a**). This is consistent with
153 iNs at this age possessing several characteristics of progenitor cells such as high *SOX2*
154 expression (**Fig. 1a and 1b**). Genes participating in cell type-specific interactions are
155 enriched for biological processes associated with their respective cell types, including cell
156 proliferation for RG and IPCs, neuron projection development for IPCs and eNs, and
157 synaptogenesis for eNs (**Supplementary Fig. 4a and Supplementary Table 3**). In
158 addition, based on comparing interaction strength and gene expression side-by-side, we
159 observe the two to be generally correlated (**Fig. 2a**). In fact, interaction strength and gene
160 expression are globally correlated across all pairwise comparisons of cell types (**Fig. 2b**
161 **and Supplementary Fig. 4b**), suggesting that gene expression is orchestrated by
162 physical chromatin looping in a manner that is highly cell type-specific.

163

164 To investigate how chromatin interactions contribute to gene regulation in greater detail,
165 we took advantage of the enrichment of open chromatin regions at distal interacting
166 regions (**Fig. 2c and Supplementary Fig. 4c**) and performed transcription factor (TF)
167 motif enrichment analysis using HOMER[14] at the set of cell type-specific distal interacting
168 regions for each cell type (**Fig. 2d and Supplementary Table 4**). PAX6, EOMES, and
169 TBR1 are the most highly enriched motifs in RG, IPCs, and eNs, respectively,
170 recapitulating their sequential expression along this developmental axis[15]. Meanwhile,

171   motifs for progenitor-specific TFs including EMX1, EMX2, and LHX2 are enriched in RG

172   and IPCs. The motif for RFX4, which was previously identified as an RG marker in the

173   murine midbrain as well as the human telencephalon, is enriched in RGs[1,16]. Finally, the

174   DLX1, DLX2, DLX6, GSX2, and LHX6 motifs are enriched in iNs, consistent with their

175   roles in iN maturation and function[17-19]. Overall, our approach identifies both known and

176   novel associations between TF binding at distal interacting regions and processes linked

177   to cellular identity.

178

179   **Super interactive promoters are enriched for lineage-specific genes.**

180

181   The number of chromatin interactions at H3K4me3-marked promoters is only modestly

182   correlated with gene expression (**Supplementary Fig. 5a**). One explanation is that

183   individual genes are expressed to varying degrees in the context of their diverse cellular

184   functions, and regulatory elements are better described as fine-tuning rather than

185   independently inducing or silencing the expression of their cognate promoters. Multiple

186   regulatory interactions can also exert synergistic or nonlinear effects on gene expression.

187   To examine the relationship between gene expression and chromatin interactivity in

188   greater detail, we first demonstrate that cell type-specific genes tend to have more

189   interactions than shared genes across all cell types (**Fig. 3a and Supplementary Fig.**

190   **5b**). Next, by ranking promoter-containing anchor bins according to their cumulative

191   interaction scores, we identify a subset of promoters with unusually high degrees of

192   chromatin interactivity, which we term super interactive promoters (SIPs) (**Fig. 3b**). In

193   total, we annotate 755, 765, 638, and 663 SIPs in RG, IPCs, eNs, and iNs, respectively

194   (**Fig. 3c and Supplementary Table 5**). SIPs are enriched for key lineage-specific genes

195   including *GFAP* and *HES1* for RG, *EOMES* for IPCs, *SATB2* for eNs, and *GAD1*, *GAD2*,

196   *DLX5*, *DLX6*, and *LHX6* for iNs. SIPs are also frequently shared across multiple cell types.

197   For example, we identify SIPs for *FOXG1* and *POU3F3* (*BRN1*) in all four cell types,

198   *SOX2* in the progenitor-like RG, IPCs, and iNs, and *TBR1* in the eN-like IPCs and eNs.

199   Interestingly, a large number of promoters for lincRNA genes including *LINC00461* and

200   *LINC01551* are annotated as SIPs, consistent with their patterns of expression in the

201   developing cortex[20]. Globally, SIPs are enriched for in cell types with the highest

202 expression of their genes among all four cell types, supporting their putative roles in
203 lineage specification (**Fig. 3d**). To assess whether SIPs are a general feature for other
204 cell types, we expanded our analysis to hematopoietic lineages with published promoter
205 capture Hi-C datasets[21]. Consistent with our results in brain cells, SIPs identified in
206 neutrophils, naive CD4+ T cells, monocytes, megakaryocytes, and erythroblasts are also
207 enriched for cell type-specific over shared genes (**Fig. 3e**). Based on these lines of
208 evidence, SIPs may represent a general mechanism used by cells to maintain the precise
209 and robust expression of key genes underlying cellular identity and function.

210

211 **Specific families of transposable elements are implicated in SIP formation.**

212

213 Given the important roles SIPs may harbor in establishing cellular identity, we were
214 interested in exploring potential mechanisms underlying their formation and evolution.
215 Towards this goal, we evaluated the contribution of transposable elements (TEs), which
216 are capable of propagating regulatory elements across the genome and influencing 3D
217 genome architecture[22-24]. First, we analyzed the enrichment of TEs at the class, family,
218 and subfamily levels in sequences defined by the union of SIPs and their distal interacting
219 regions (SIP groups or SIPGs) (**Fig. 3f and Supplementary Fig. 6a-c**). Notably, the
220 ERVL-MaLR family and many of its subfamilies are enriched in SIPGs across all four cell
221 types. Since we detected the strongest enrichment of this family of TEs in eNs, we
222 decided to focus on this particular lineage. In total, we identified 16 SIPGs in eNs that are
223 statistically enriched for ERVL-MaLR TEs (hypergeometric test, $P < 0.01$) (**Fig. 3g**). Next,
224 we used HOMER to perform TF motif enrichment analysis at ERVL-MaLR TEs within
225 these 16 SIPGs and determined ZNF143 to be the most enriched motif (**Fig. 3h**). ZNF143
226 is an architectural protein which has previously been reported to mediate looping between
227 promoters and distal regulatory elements[25]. Moreover, certain subfamilies of ERVL-MaLR
228 TEs have been demonstrated to contribute to ZNF143 binding in 3T3 and HeLa cells[26].
229 The *ADRA2A* SIPG in eNs exhibited the highest enrichment of ERVL-MaLR-derived
230 ZNF143 motifs (hypergeometric test, $P=1.59 \times 10^{-6}$) (**Fig. 3i**) and was associated with
231 elevated *ADRA2A* expression in eNs (**Supplementary Fig. 6d**). It spans 42 distal
232 interacting regions, 25 of which contain ERVL-MaLR TEs, and 12 of which contain one

233   or more ERVL-MaLR-derived ZNF143 motifs (**Fig. 3j and Supplementary Fig. 6e**).

234   Furthermore, ZNF143 motifs in TEs from multiple ERVL-MaLR subfamilies (THE1A,

235   THE1C, MSTA) within the SIPG can be mapped back to ZNF143 motifs in the consensus

236   sequences of the same subfamilies (**Supplementary Fig. 6f**). This supports a model in

237   which ZNF143 motifs are coordinately expanded by ERVL-MaLR TE insertion, leading to

238   increased binding site redundancy and strengthened assembly of the *ADRA2A* SIPG

239   regulatory unit (**Fig. 3k**). Our results imply that TEs are capable of mediating the formation

240   of higher order epigenomic features including SIPs[27].

241

242   **Investigating features of the developmental trajectory from RG to eNs during**

243   **corticogenesis.**

244

245   Since RG, IPCs, and eNs represent a developmental trajectory from dorsal cortex

246   progenitors to mature functional neurons, we grouped genes according to their

247   expression and cumulative interaction scores along this axis and identified groups

248   corresponding to RG, IPCs, and eNs (groups 1-3) that are enriched for lineage-defining

249   genes and biological processes (**Fig. 4a, Supplementary Fig. 7, and Supplementary**

250   **Table 6**). We also identified groups with decreasing expression and increasing chromatin

251   interactivity (group 4) as well as increasing expression and decreasing chromatin

252   interactivity (group 5) from RG to eNs, which could represent late-silenced and early-

253   silenced genes, respectively. Late-silenced genes are enriched for chromatin remodeling

254   and epigenetic regulation terms, whereas early-silenced genes are enriched for eN-

255   specific signatures. These results demonstrate that gene expression can be mediated by

256   distinct modes of chromatin-mediated regulation during development.

257

258   Human corticogenesis is dramatically distinct from that in other mammalian species,

259   driven in large part by the increased diversity and proliferative capacity of cortical

260   progenitors during development which results in the increased size and complexity of the

261   human brain[28]. In particular, Notch signaling genes have been implicated in the clonal

262   expansion of RG, which constitute the major subtype of cortical progenitors in the

263   cortex[29,30]. Here, we find that RG are enriched relative to other cell types for chromatin

264   interactions at Notch signaling genes from the AmiGO database[31] (**Fig. 4b**). Compared

265   to other cell types, chromatin interactions in RG also target a significantly higher

266   proportion of human gained enhancers identified through comparative analyses of

267   H3K4me2 and H3K27ac ChIP-seq signal in human, rhesus macaque, and mouse

268   brains[32]. Therefore, 3D epigenomic landscapes are capable of identifying lineage-specific

269   pathways contributing to human-specific aspects of cortical development. In addition, we

270   provide detailed annotations of gene targets for human gained enhancers and in vivo-

271   validated enhancer elements from the Vista Enhancer Browser[33] in **Supplementary**

272   **Table 7**.

273

274   **Leveraging 3D epigenomic landscapes to partition heritability for complex**

275   **neuropsychiatric disorder- and trait-associated variants.**

276

277   Chromatin interactions identified in our sorted cell populations represent a unique in vivo

278   resource for mapping complex neuropsychiatric disorder- and trait-associated variants to

279   their target genes (**Figure 4c and Supplementary Table 8**). They additionally enable the

280   assessment of cell type-specific patterns of SNP heritability. To partition SNP heritability

281   using our 3D epigenomic annotations, we employed linkage disequilibrium score

282   regression (LDSC)[34,35] using summary statistics from genome wide association studies

283   (GWAS) for the following neuropsychiatric traits: Alzheimer's disease (AD), attention

284   deficit hyperactivity disorder (ADHD)[36], autism spectrum disorder (ASD)[37], bipolar

285   disorder (BD)[38], intelligence quotient (IQ)[39], major depressive disorder (MDD)[40], and

286   schizophrenia (SCZ)[41]. Overall, we observed significant levels of heritability enrichment

287   in 3D anchor bins for every cell type and neuropsychiatric trait we evaluated ($15.13 <$

288   enrichment score < 51.57, $1.27 \times 10^{-40} < P < 0.02$) (**Fig. 4d**). These findings are largely

289   expected as the majority of interacting promoters are shared across our cell types. When

290   we restricted our analysis to 3D target bins, we observed dramatically distinct patterns of

291   cell type-specific heritability enrichment (**Fig. 4e**). For example, ASD SNP heritability was

292   significantly enriched for only in RG and eNs ($P = 1.07 \times 10^{-3}$ and 4.75×10^{-4}, respectively),

293   and AD SNP heritability was significantly enriched for only in RG ($P = 3.99 \times 10^{-6}$). Our

294   findings reflect the cell type-specific nature of distal regulatory elements that are

295   dysregulated during disease and underscore the importance of leveraging 3D epigenomic

296   annotations to interpret variants that are located in non-coding regions of the genome.

297

298   **Functional characterization of enhancers in primary cells using CRISPRview.**

299

300   Validating distal regulatory elements in primary cells has historically been challenging,

301   with most experiments to date performed using cell lines or iPSC-derived cells[10,42]. Here,

302   we present CRISPRview, a novel approach combining CRISPRi, RNAscope, and

303   immunostaining to validate enhancers in heterogeneous cultures of primary cells at the

304   single cell level (**Fig. 5a**). We use CRISPRview to validate multiple enhancers in RG at

305   the *GPX3*, *TNC*, and *HES1* loci, all of which harbor RG-specific chromatin interactions

306   and are differentially expressed in RG (**Fig. 5b-d**). Furthermore, the promoters for *TNC*

307   and *HES1* are annotated as SIPs in RG. First, sgRNAs were designed to target open

308   chromatin regions physically interacting with the promoters of *GPX3*, *TNC*, and *HES1*

309   (**Supplementary Table 9**). Next, primary cultures of microdissected and dissociated GZ

310   samples between GW17 to GW19 were infected with lentivirus expressing the

311   experimental sgRNA, dCas9-KRAB, and mCherry in combination with lentivirus

312   expressing control sgRNA, dCas9-KRAB, and GFP. After five additional days in culture,

313   the cells were fixed and stained using antibodies for mCherry, GFP, the RG marker

314   GFAP, and RNAscope probes (ACD[TM]) targeting intronic regions of the genes of interest.

315   Finally, high resolution images were taken using confocal microscopy, and the number of

316   punctate dots representing individual nascent transcripts were compared between

317   experimental (mCherry+) and control (GFP+) sgRNA-treated GFAP+ RG (**Fig. 5b-d**). The

318   SMART-Q pipeline was specifically developed in our lab for image analysis (see attached

319   manuscript, submitted).

320

321   All four regions interacting with the *GPX3* promoter (regions 1-4) were found to exhibit

322   downregulation of *GPX3* expression upon CRISPRi targeting (**Fig. 5b**). Notably, region 1

323   overlaps both a human gained enhancer[32] and a Vista enhancer element (mm1343)[33],

324   supporting its function as an enhancer in RG. Next, we investigated the locus for *TNC*, a

325   RG-specific gene implicated in neuronal migration, axon guidance, and synaptic plasticity.

326 We found that two of its interacting regions exhibited significant downregulation of *TNC*
327 expression upon CRISPRi targeting (regions 1 and 2), but that silencing of its other two
328 interacting regions did not result in notable changes in *TNC* expression (regions 3 and 4)
329 (**Fig. 5c**). This could be due to the presence of alternative regulatory elements or
330 structural interactions at these loci. Finally, HES1 is a lineage-defining TF for RG, and we
331 positively validated all three regions that were found to interact with the *HES1* promoter
332 (regions 1-3) (**Fig. 5d**). The observation of small but significant changes in gene
333 expression upon CRISPRi targeting supports the hypothesis that multiple interactions
334 work in concert to titrate the expression of key genes linked to cellular identity. In addition,
335 the observed broad distributions of nascent transcript counts likely reflects the stochastic
336 nature of transcription in single cells, stressing the importance of employing an approach
337 combining resolution, sensitivity, and cell type-specificity for validating enhancers in
338 single cells.

339

340 **Discussion**

341

342 Recent publications leveraging single cell sequencing have highlighted the heterogeneity
343 of the developing human cortex, underscoring the necessity of studying epigenomic
344 regulation in a cell type-specific manner. Within the dorsal cortex alone, there are a
345 massive variety of cell types from multiple sources both within and outside the developing
346 neural tube, including RG, IPCs, eNs, MGE-derived iNs, CGE-derived iNs, microglia,
347 endothelial cells, and subplate neurons. Despite large differences in maturation state and
348 lineage, many of these cell types share intriguing similarities in terms of gene expression.
349 For example, iNs express a number of TFs typically associated with RG proliferation such
350 as *SOX2* or eN differentiation such as *ASCL1* and *NPAS3*[1]. Therefore, bulk
351 measurements cannot reliably reveal the nuanced epigenomic programs driving gene
352 expression in each cell type. By profiling 3D epigenomic landscapes in specific cell
353 populations during corticogenesis, we not only demonstrate that gene regulation is closely
354 linked to chromatin interactivity, we also identify SIPs that are highly cell type-specific and
355 enriched for key lineage-specific genes. We uncover a potential mechanism by which
356 specific families of TEs propagate binding sites for architectural proteins, facilitating the

357 formation of multi-interaction clusters which may serve to sustain gene expression. While
358 the analysis of TEs is currently constrained by the list of known motifs and the resolution
359 of chromatin interactions identified in this study, future advances will help us further
360 elucidate the contribution of TEs to 3D genome architecture and transcriptional regulation.
361

362 Cortical progenitors, eNs, and iNs are highly divergent in terms of their diversity,
363 proliferative capacity, distribution, and functional characteristics between humans and
364 mice[28]. Therefore, processes occurring during human cortex development cannot be fully
365 recapitulated with mouse models. These non-murine features also indicate that enhancer
366 mutations in humans may not adequately phenocopy to mice. Our dataset provides a
367 comprehensive catalog of annotations for human gained enhancers and complex
368 neuropsychiatric disorder- or trait-associated variants in cell types that are intricately tied
369 to human cortex development, enabling the interpretation and prioritization of regulatory
370 sequences for follow-up studies.
371

372 Finally, there is a need to perform cell type-specific validation for regulatory sequences in
373 primary cells, especially as our understanding of epigenomic regulation matures over
374 time. By combining immunostaining for cellular markers with the quantification of nascent
375 transcripts in the nucleus, CRISPRview offers exquisite sensitivity and resolution for
376 detecting cell type-specific changes in gene expression in single cells. Here, we used
377 CRISPRview to successfully validate multiple regulatory elements in RG and observed
378 subtle but significant changes in gene expression at the *GPX3*, *TNC*, and *HES1* loci.
379 Further experiments leveraging CRISPRview in live tissue cultures should continue to
380 reveal novel regulatory logic in a manner that is truly representative of the complex in vivo
381 environment that is present during human cortex development.
382

383 **Acknowledgements**
384

398

399   **Author contributions**

400

401   Y.S., A.A.P., and A.R.K. conceived the study. Y.S., M.H., A.A.P., A.R.K. supervised the
402   study. M.S., M.P., X.Y., I.R.J., X.C., U.C.E., and L.M. performed experiments. M.S., A.A.,
403   S.B., J.D.R., B.L., I.J., and M.H. performed computational analysis. C.F. and M.N.C.
404   performed transposable element analysis under the supervision of T.W. J.W. and W.L.
405   performed SNP heritability analysis under the supervision of Y.L. M.S., M.P., X.Y., and
406   Y.S. analyzed and interpreted the data. Y.S., M.S., M.P., X.Y., and M.H prepared the
407   manuscript with input from all other authors.

408

409   **Competing interests statement**

410

411   The authors declare no competing financial interests.

412

413   **Code availability statement**

414

415   A copy of the custom code used for data analysis and figure generation in this study is
416   available upon request.

417

418   **Data availability statement**

419

420   All datasets used in this study (PLAC-seq, ATAC-seq, RNA-seq) are available at the

421   Neuroscience Multi-Omic Archive (NeMO Archive) under controlled access. Chromatin

422   interactions, open chromatin regions, and gene expression results for each cell type can

423   be accessed from the NeMO Archive using the following link:

424   https://assets.nemoarchive.org/dat-uioqy8b

425

426   Data can also be visualized on the WashU Epigenome Browser using the following link:

427   http://epigenomegateway.wustl.edu/legacy/?genome=hg38&session=OCzw03b5lz&stat

428   usId=1958712809

429

430   **Figure legends**

431

432   **Figure 1. Experimental design and general features of the 3D epigenomic**

433   **landscape during corticogenesis.**

434

435   (**a**) Schematic of the sorting strategy. Within the dorsal cortex, the germinal zone (GZ) is

436   populated by radial glia (RG), which extend fibers towards the cortical plate (CP). These

437   RG divide asymmetrically to produce intermediate progenitor cells (IPCs), which

438   differentiate into excitatory neurons (eNs) that migrate along RG fibers towards the CP.

439   At the same time, interneurons (iNs) can be found in both the GZ and CP. Microdissected

440   GZ and CP samples were dissociated into single cells before being fixed, stained with

441   antibodies for EOMES, SOX2, PAX6, and SATB2, and sorted using FACS. (**b**) Heatmap

442   showing expression of key marker genes for RG, IPCs, eNs, and iNs. (**c**) Heatmap

443   showing correlations between gene expression profiles for sorted cell populations and

444   aggregate gene expression profiles from scRNA-seq datasets in the developing cortex.

445   Cell types include newborn iNs from the medial ganglionic eminence (MGE), ventral

446   progenitors including RG and IPCs from the MGE, microglia, and choroid plexus cells. (**d**)

447   WashU Epigenome Browser snapshot of a 360 kb region (chr17:72,970,0000-

448   73,330,000) showing IPC-specific chromatin interactions linked to *SSTR2* expression in

449   IPCs. (e) Bar graph showing counts of MAPS interactions, with proportions of XOR (blue,

450   only one interacting bin contains H3K4me3 peaks) and AND (red, both interacting bins

451   contain H3K4me3 peaks) interactions displayed for each cell type. (**f**) Cumulative

452   distribution function (CDF) plots showing interaction distances for each cell type. (**g**)

453   Histogram showing the numbers of MAPS interactions at each promoter for each cell

454   type.

455

456   **Figure 2. H3K4me3-mediated chromatin interactions contribute to cell type-specific**

457   **gene regulation.**

458

459   (**a**) Heatmaps displaying interaction scores (left) and gene expression (right) for unique

460   XOR interactions grouped according to their cell type specificity. Hierarchical clustering

461   dendrograms for each heatmap are also shown (top). (**b**) Scatterplot showing positive

462   correlation between the difference in the number of MAPS interactions at each promoter

463   and the difference in expression of the corresponding genes between RG and eNs

464   (Pearson product-moment correlation coefficient, two-sided t-test, $P < 2.2 \times 10^{-16}$). The

465   fitted trendline based on linear regression is also shown. (**c**) Fold enrichment of open

466   chromatin regions over distance-matched background regions in 1 Mb windows around

467   distal interacting regions for MAPS interactions in RG. (**d**) Enrichment of TF motifs at

468   open chromatin regions in cell type-specific interacting distal regions for each cell type.

469   The color of each dot represents the degree of enrichment ($-\log_{10}$P-value) for each TF

470   motif, and the size of each dot represents the gene expression of the corresponding TF.

471

472   **Figure 3. Super interactive promoters are enriched for lineage-specific genes.**

473

474   (**a**) CDF plots showing the numbers of MAPS interactions for shared versus cell type-

475   specific genes in eNs (two sample t-test, two-sided, $P = 1.40 \times 10^{-14}$). (**b**) Plots showing

476   the ranked cumulative interaction scores for 3D anchor bins in each cell type, defined as

477   the sum of the $-\log_{10}$FDR for MAPS interactions coincident to each bin. Super interactive

478   promoters (SIPs) are defined as promoters located to the right of the knee of each curve

479   (dashed lines). Example SIPs, including those for lineage-specific genes, are highlighted

480   for each cell type. (**c**) Venn diagram displaying the cell type-specificity of SIPs in RG,

481   IPCs, eNs, and iNs. (**d**) The number of genes called as SIPs was divided by the total

482   number of SIPs and non-SIPs for genes with the 1st, 2nd, 3rd, and 4th highest expression

483   among all four cell types. The fold enrichment was calculated relative to the group with

484   the 4th highest expression for each cell type. (**e**) Forrest plot showing that SIPs called in

485   hematopoietic cells are also enriched for cell type-specific over shared genes. 95%

486   confidence intervals are displayed. (**f**) Scatterplot showing both the enrichment and the

487   number of observed copies of TE families in SIPGs for eNs. TE families occupying more

488   than 1% of the genome are colored. (**g**) Scatterplot showing the enrichment of ERVL-

489   MaLR TEs in SIPGs for eNs (hypergeometric P-value, see methods). SIPGs with 40 or

490   more distal interacting bins and $P < 0.01$ are highlighted. (**h**) Scatterplot showing the

491   enrichment of TF motifs at ERVL-MaLR TEs in SIPGs highlighted in (**g**). Only TF motifs

492   with length > 12 bp are shown. (**i**) Scatterplot showing the enrichment of ZNF143 motifs

493   at ERVL-MaLR TEs in SIPGs highlighted in (**g**) (Poisson distribution, see methods).

494   ZNF143 motifs occurrences were detected using FIMO using a threshold of $P = 0.0001$.

495   (**j**) WashU Epigenome Browser snapshot of the *ADRA2A* SIPG. MAPS interactions

496   targeting the 12 distal interacting bins containing ERVL-MaLR-derived ZNF143 motifs are

497   highlighted. (**k**) Potential mechanism for the contributions of TEs towards SIP formation.

498

499   **Figure 4. Investigating developmental trajectories during corticogenesis and**

500   **partitioning heritability for complex neuropsychiatric disorders and traits.**

501

502   (**a**) Gene groups identified based on their changes in expression and chromatin

503   interactivity along the transition from RG to eNs. Group 1 represents stem cell genes with

504   decreasing expression and chromatin interactivity from RG to eNs. Group 2 represents

505   IPC-specific genes with the highest expression and chromatin interactivity at the IPC

506   stage. Group 3 represents genes with increasing expression and chromatin interactivity

507   from RG to eNs. Groups 4 and 5 are characterized by anti-correlated expression and

508   chromatin interactivity and may represent late-silenced and early-silenced genes,

509   respectively. Representative genes and GO terms are shown for each group. (**b**) Bar

510   graph showing the numbers of MAPS interactions at Notch signaling genes targeting bins

511   with and without human gained enhancers in each cell type (Chi-square test). (**c**) Bar

512   graph showing the numbers of unique GWAS SNPs ($P < 10^{-8}$) interacting with their

513   nearest gene only, with both their nearest and distal genes, or with distal genes only for

514   each cell type across all neuropsychiatric traits. (**d-e**) LDSC enrichment scores for each

515   neuropsychiatric trait and cell type, stratified by 3D anchor and target bins. Results with

516    $P > 0.05$ are indicated.

517

518   **Figure 5. Functional characterization of distal interacting regions using**

519   **CRISPRview.**

520

521   (**a**) CRISPRview workflow. Image analysis was performed using the SMART-Q pipeline.

522   (**b-d**) Functional characterization of distal interacting regions at the *GPX3*, *TNC*, and

523   *HES1* loci. For each locus, a WashU Epigenome Browser snapshot shows chromatin

524   interaction bridging the promoters of *GPX3*, *TNC*, and *HES1* and distal interacting regions

525   containing open chromatin regions (highlighted) which were targeted by sgRNAs for

526   CRISPRi silencing. Representative images show staining for RNAscope probes targeting

527   intronic regions for the genes of interest (white), DAPI (blue), the RG marker GFAP (light

528   blue), mCherry (red), and GFP (green). The scale bar is 50 μm. Box plots show the results

529   of CRISPRi silencing for each targeted region. The open circles represent single cells,

530   and nascent transcript counts for experimental (mCherry+) versus control (GFP+)

531   sgRNA-treated RG are represented on the y-axis (Student's t-test, two-tailed). The

532   median, upper and lower quartiles, and 10% to 90% range are indicated.

533

534   **Supplemental Figure 1. Representative contour plots depicting FACS gating**

535   **strategy.**

536

537   (**a**) Cells were separated from debris of various sizes based on the forward scatter area

538   (FSC-A) and side scatter area (SSC-A). Cells were then passed through two singlet gates

539   using the width and height metrics of the (**b**) side scatter (SSC-H versus SSC-W) and (**c**)

540   forward scatter (FSC-H versus FSC-W). (**d**) SOX2+, and SOX2-, and intermediate

541   progenitor (IPC) populations were isolated by gating on EOMES-PE-Cy7 and SOX2-

542   PerCP-Cy5.5 staining. (**e**) Radial glia (RG) and interneurons (iNs) were isolated as high

543   PAX6/high SOX2 and medium SOX2/low PAX6 populations, respectively. (**f**) Excitatory

544   neurons (eNs) were isolated from the SOX2- population by gating on SATB2-Alexa Fluor

545   647 staining.

546

547   **Supplementary Figure 2. Reproducibility between replicates for RNA-seq, ATAC-**

548   **seq, and PLAC-seq.**

549

550   (**a**) RNA-seq replicates were hierarchically clustered according to gene expression

551   sample distances using DESeq2. (**b**) Heatmap with pairwise correlations and hierarchical

552   clustering for read densities at the set of unified open chromatin regions for ATAC-seq

553   replicates. (**c**) Principle component analysis (PCA) was performed based on the

554   normalized contact frequencies across all PLAC-seq replicates (see methods). To assess

555   the robustness of the results, we conducted the analysis separately for bin pairs within

556   300 and 600 kb interacting windows.

557

558   **Supplementary Figure 3. Identification of chromatin interactions using MAPS.**

559

560   (**a**) Illustration of AND and XOR sets in a representative PLAC-seq contact matrix. The

561   blue tracks represent 1D H3K4me3 peaks at bin positions 4, 8, 12, and 14. The black

562   cells represent interactions within the same bin. The purple cells represent interactions in

563   the AND set where both of the interacting bins contain 1D H3K4me3 peaks. The orange

564   cells represent interactions in the XOR set where only one of the interacting bins contains

565   1D H3K4me3 peaks. The grey cells represent interactions where neither of the interacting

566   bins contains 1D H3K4me3 peaks. (**b**) Venn diagram displaying cell type-specificity of

567   MAPS interactions for each cell type. (**c**) Proportions of MAPS interactions occurring

568   within and across TADs in GZ and CP tissues for each cell type.

569

570   **Supplementary Figure 4. Contribution of 3D epigenomic landscapes to gene**

571   **regulation.**

572

573   (**a**) GO enrichment analysis for genes whose promoters participate in cell type-specific

574   interactions. The top annotation clusters from DAVID are reported along with their

575   enrichment scores for each cell type. (**b**) Scatterplots showing positive correlation

576   between the difference in the number of MAPS interactions at each promoter and the

577   difference in expression of the corresponding genes between all pairs of cell types

578   (Pearson product-moment correlation coefficient, two-sided t-test, $P < 2.2 \times 10^{-16}$ for all cell

579   types). Fitted trendlines based on linear regression are also shown. (**c**) Fold enrichment

580   of open chromatin regions over distance-matched background regions in 1 Mb windows

581   around distal interacting regions for MAPS interactions in IPCs, eNs, and iNs.

582

583   **Supplementary Figure 5. Correlations between chromatin interactions and gene**

584   **expression for cell-type specific and shared genes.**

585

586   (**a**) Scatterplots showing the correlation between numbers of MAPS interactions and gene

587   expression at promoters in each cell type. (**b**) Cumulative distribution function (CDF) plots

588   showing the numbers of MAPS interactions for shared versus cell type-specific genes in

589   RG, IPCs, and iNs (two sample t-test, two-sided, $P = 5.91 \times 10^{-22}$, 3.74×10^{-11}, and 2.94×10^{-19}

590   for RG, IPCs, and iNs, respectively).

591

592   **Supplementary Figure 6. Specific families of transposable elements are implicated**

593   **in SIP formation.**

594

595   (**a-c**) Enrichment of TEs in SIPGs at the class (**a**), family (**b**), and subfamily (**c**) levels for

596   each cell type. Only families occupying more than 1% of the genome are shown in (**b**).

597   Only subfamilies from the MIR and ERVL-MaLR families occupying more than 0.1% of

598   the genome are shown in (**c**). (**d**) Bar graph shows elevated *ADRA2A* gene expression in

599   eNs. (**e**) Illustration of the 12 distal interacting regions in the *ADRA2A* SIPG containing at

600   least one ERVL-MaLR-derived ZNF143 motif. ZNF143 motifs are indicated and colored

601   by strand. The bin identifier corresponds to the labels in **Fig. 3j**. (**f**) Illustration of the

602   conservation of ZNF143 binding motifs in ERVL-MaLR TEs. Blue bars indicate consensus

603   sequences, yellow bars indicate individual copies of ERVL-MaLR TEs in the *ADRA2A*

604   SIPG, and red bars indicate ZNF143 motifs. The positions of the ZNF143 motifs relative

605   to the ERVL-MaLR TE sequences was determined using FIMO.

606

607   **Supplementary Figure 7. Developmental trajectories and annotations for complex**

608   **neuropsychiatric disorder- and trait-associated variants.**

609

610   (**a**) Box plots showing the distributions of gene expression and cumulative interaction

611   scores for groups in **Fig. 4**. The median, upper and lower quartiles, minimum, and

612   maximum are indicated. (**b**) Bar graphs showing the numbers of GWAS SNPs ($P < 10^{-8}$)

613   interacting with their nearest gene only, with both their nearest and distal genes, or with

614   distal genes only for each cell type and neuropsychiatric trait. Venn diagrams display the

615   cell type-specificity of all interacting GWAS SNPs for each neuropsychiatric trait.

616

617   **Supplementary Table 1. Sample metadata.**

618

619   **Supplementary Table 2. PLAC-seq, ATAC-seq, and RNA-seq data processing**

620   **metrics.**

621

622   **Supplementary Table 3. Enriched GO terms for genes participating in cell type-**

623   **specific interactions.**

624

625   **Supplementary Table 4. Motif enrichment at cell type-specific distal interacting**

626   **regions.**

627

628   **Supplementary Table 5. Super interactive promoters for each cell type.**

629

630   **Supplementary Table 6. Enriched GO terms for genes associated with specific**

631   **developmental trajectories.**

632

633   **Supplementary Table 7. Target gene annotation for enhancers overlapping**

634   **chromatin interactions.**

635

Supplementary Table 8. Target gene annotation for complex neuropsychiatric disorder- and trait-associated GWAS SNPs overlapping chromatin interactions.

638

Supplementary Table 9. sgRNA sequences used for functional validation.

640

Methods

642

Ethics statement

644

Deidentified embryonic brain tissue samples were collected with prior patient consent in strict observance of legal and institutional ethical regulations. All protocols were approved by the Human Gamete, Embryo, and Stem Cell Research Committee (GESCR) and Institutional Review Board at the University of California, San Francisco.

649

Tissue dissociation

651

The tissue dissociation protocol was adapted from Nowakowski et al, 2017[1]. Briefly, samples were cut into small pieces in artificial cerebrospinal fluid before being added into pre-warmed papain dissociation media (Worthington #LK003150). The dissociation solution was incubated for 45 minutes at 37°C. Excess solution was removed and replaced with cell culture media. The pieces of tissue were triturated, filtered through a 70 μM nylon mesh, and centrifuged at 300 g for 8 minutes. The supernatant was removed and replaced with fresh culture media.

659

Sample fixation

661

Mid-gestational human brain samples between GW15 and GW22 were fixed in 2% paraformaldehyde prepared in PBS with gentle agitation for 10 minutes at room temperature. Glycine was added to a final concentration of 200 mM to quench the

665 reactions, and the samples were centrifuged at 500 g for 5 minutes at 4°C. The samples

666 were washed twice with PBS before being frozen at -80°C until further processing.

667

668 **Permeabilization and staining**

669

670 The cell pellet was thawed on ice and resuspended in PBS with 0.1% Triton-X-100 for 15

671 minutes. The cells were then washed twice with PBS and resuspended in 5% BSA in PBS

672 for staining. Staining proceeded for at least one hour with FcR Blocking Reagent (Miltenyi

673 Biotech, 1/20 dilution), EOMES-PE-Cy7 (Invitrogen, WD1928, 1/10 dilution), PAX6-PE

674 (BD Biosciences, O18-1330, 1/10 dilution), SOX2-PerCP-Cy5.5 (BD Biosciences, O38-

675 678, 1/10 dilution), and SATB2-Alexa Fluor 647 (Abcam, EPNCIR130A, 1/100 dilution).

676 Following staining, the cells were centrifuged at 500 g for 5 minutes. The supernatant was

677 removed, and the pellet was diluted into PBS. When sorting cells for RNA-seq libraries,

678 1% RNasin Plus RNase Inhibitor (Promega) was included in all buffers, PBS was

679 prepared from RNase-free stocks, and acetylated RNase-free BSA was used to prepare

680 5% BSA in PBS for staining.

681

682 **FACS**

683

684 AbC Total Antibody Compensation Beads (Thermo Fisher) were used to generate single

685 color compensation controls prior to sorting. Sorting was conducted on either the

686 FACSAria II, FACSAria IIu, or FACSAria Fusion instruments using a 70 µM nozzle, and

687 cells were collected in 5 ml tubes pre-coated with FBS. A sample of each sorted

688 population was reanalyzed on the same machine to assess purity. The cells were

689 collected by centrifuging at 500 g for 10 minutes. The supernatant was removed, and the

690 pellet was frozen at -80°C until further processing. When sorting cells for RNA-seq

691 libraries, collection tubes were coated with both FBS and RNAlater (Thermo Fisher)

692

693 **Primary cell culture**

694

695 Following dissociation, cells were plated onto Matrigel-coated coverslips in 48 well plates

696   at a density of approximately 0.7×10^6 cells per well. The cells were infected with lentivirus

697   24 hours after plating, and media was changed every two days. Media was composed of

698   96% DMEM/F-12 with GlutaMAX, 1% N-2, 1% B-27, and 1% penicillin/streptomycin. The

699   cells were grown in 8% oxygen and 5% carbon dioxide, and they were harvested for

700   fixation four days post-infection.

701

702   **PLAC-seq**

703

704   PLAC-seq was performed according to the protocol from Fang et al., 2016[43]. 1 to 5 million

705   cells were used to prepare each library. Digestion was performed using 100 U MboI for 2

706   hours at 37°C, and chromatin immunoprecipitation was performed using Dynabeads M-

707   280 sheep anti-rabbit IgG (Invitrogen #11203D) mixed with 5 µg anti-H3K4me3 antibody

708   (Millipore 04-745). TruSeq sequencing adapters were added during PCR amplification.

709   Libraries were sent for paired-end sequencing on the HiSeq X Ten or NovoSeq 6000

710   instruments (150 bp paired-end reads). fastp was applied to trim reads to 100 bp, and

711   replicates were merged and downsampled to normalize the number of usable reads

712   before processing with MAPS.

713

714   **MAPS interaction calling**

715

716   We used the MAPS pipeline to call significant long-range chromatin interactions from our

717   PLAC-seq datasets. First, bwa mem was used to map raw reads to hg38. Unmapped

718   reads and reads with low mapping quality were discarded, and the resulting filtered read

719   pairs were processed as previously reported[13]. Briefly, we divided the genome into 5 kb

720   bins and counted the number of read pairs representing interactions between 5 kb bins.

721   To define our 3D anchor bins, we took the union of peaks called using MACS2 from read

722   pairs with interaction distances < 1 kb for each cell type (1D H3K4me3 peaks). Based on

723   this annotation, we classified interactions into AND, XOR, and NOT sets based on

724   whether both, only one, or none of the interacting 5 kb bins overlapped 1D H3K4me3

725   peaks (**Supplementary Fig. 3a**). Since we were interested in identifying significant

726   H3K4me3-mediated chromatin interactions, we retained only interactions in the AND and

727   XOR sets for downstream processing. We also retained only intrachromosomal

728   interactions with interaction distances between 10 kb and 1 Mb. These two criteria

729   constituted our definition of usable reads.

730

731   For calling interactions, we applied a Poisson regression-based approach to normalize

732   systematic biases from restriction sites, GC content, sequence repetitiveness, and ChIP

733   enrichment. We fitted models for interactions in the AND and XOR sets separately and

734   calculated false discovery rates (FDRs) for interactions based on their expected and

735   observed contact frequencies between 5 kb bin pairs. Furthermore, we grouped

736   interactions that were located within 15 kb of each other at both ends into clusters and

737   classified all other interactions as singletons. To define our sets of significant long-range

738   chromatin interactions, we retained only interactions with 12 or more reads, normalized

739   contact frequencies (defined as the ratio between the observed and expected contact

740   frequencies) ≥ 2, and FDR < 0.01 for clusters and FDR < 0.0001 for singletons. This was

741   based on the reasoning that more biologically meaningful interactions are likely to appear

742   in clusters, and singletons are more likely to present false positives.

743

744   **MAPS reproducibility analysis**

745

746   PCA plots were generated based on the normalized contact frequencies for 5 kb bin pairs

747   from our PLAC-seq datasets. Specifically, we first extracted AND and XOR 5 kb bin pairs

748   based on cell type-specific 1D H3K4me3 peaks for each of the 11 replicates. We next

749   applied zero-truncated Poisson regression, adjusting for the same systematic biases as

750   the MAPS pipeline. Again, we derived normalized contact frequencies based on the ratio

751   between the observed and expected contact frequencies between 5 kb bin pairs, with the

752   expected contact frequencies being the fitted values from the zero-truncated Poisson

753   regression. Normalized contact frequencies were then log-transformed and merged

754   across the 11 replicates. This quantile normalized merged data was used to generate the

755   PCA plots. We restricted our analysis to 5 kb bin pairs within 300 kb or 600 kb windows

756   for **Supplementary Fig. 2c**.

757

ATAC-seq

759

ATAC-seq was performed as previously described using the Nextera DNA Library Prep Kit (Illumina #FC-121-1030). First, fixed cells were washed once with ice cold PBS containing 1x protease inhibitor before being resuspended in ice cold nuclei extraction buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl$_2$, 0.1% Igepal CA630, and 1x protease inhibitor) for 5 minutes. Next, 50,000 cells were counted out, exchanged into 50 µL 1x Buffer TD, then incubated with 2.5 µL TDE1 enzyme for 45 minutes at 37°C with shaking. Following transposition, 150 µL reverse crosslinking solution (50 µL 1 M Tris pH 8.0, 100 µL 10% SDS, 2 µL 0.5 M EDTA, 10 µL 5 M NaCl, 800 µL water, and 2.5 µL 20 mg/mL Proteinase K) was added to each reaction, and the reactions were incubated at 65°C overnight. On the next day, DNA was purified using Qiagen MinElute spin columns, amplified using Nextera primers, then size-selected for fragments between 300 and 1000 bp using AMPure XP beads. Libraries were sent for paired-end sequencing on the NovaSeq 6000 instrument (150 bp paired-end reads). Raw reads were mapped to hg38 and processed using the ENCODE pipeline (https://github.com/kundajelab/atac_dnase_pipelines) running the default settings. All sequencing reads were trimmed to 50 bp prior to mapping. The sets of optimal naive overlap peaks for each cell type were used for further downstream analysis.

777

RNA-seq

779

We extracted total RNA from the sorted cell populations using the RNAstorm™ FFPE RNA extraction kit (Cell Data Sciences #CD501) starting from 5×10^5 to 1.5×10^6 cells. The quality of the extracted RNA was checked by calculating the percentage of RNA fragments with size > 200 bp (DV200) from the Agilent 2100 Bioanalyzer. RNA samples with DV200 >= 40% were used for library construction. First, they were depleted of ribosomal RNA using the KAPA RNA HyperPrep Kit with RiboErase (HMR #KK8560). Next, the RNA was used for first and second strand synthesis, dA-tailing, and sequencing adapter ligation. The cDNA was cleaned up and TruSeq sequencing adapters were added

788   via PCR amplification. Libraries were sent for paired-end sequencing on the NovaSeq

789   6000 instrument (150 bp paired-end reads). Raw reads were aligned to hg38 using STAR

790   running the standard ENCODE parameters, and transcript quantification was performed

791   in a strand-specific manner using RSEM with the GENCODE 29 annotation. The edgeR

792   package in R was used to calculate TMM-normalized RPKM values for each gene based

793   on the expected counts and gene lengths for each replicate as reported by RSEM. The

794   mean gene expression across all replicates for each cell type was used for further

795   downstream analysis.

796

797   **TF motif enrichment analysis**

798

799   We took the set of open chromatin regions participating in cell type-specific XOR

800   interactions for each cell type and used the sequences in 200 bp windows around the

801   peak summits to perform motif enrichment analysis using HOMER running the default

802   settings. The complete set of vertebrate motifs from the JASPAR database were used for

803   detection. The "-float" option was specified to optimize the detection threshold, and the

804   entire genome was used as a background. Entries with similar or identical consensus TF

805   motif sequences were grouped for brevity.

806

807   **GO enrichment analysis**

808

809   Protein coding and noncoding RNA genes from GENCODE 29 participating in cell type-

810   specific XOR interactions were used for GO enrichment analysis. Only genes participating

811   in interactions with promoter open chromatin regions on one end and distal open

812   chromatin regions on the other end were used. A minimum normalized RPKM of 0.5 was

813   used to filter out genes that were not significantly expressed the corresponding cell types,

814   and the resulting gene lists were input into DAVID 6.8 running functional annotation

815   clustering with the default settings and the "GOTERM_BP_ALL" ontology.

816

817   **SIP identification**

818

819   We devised an approach similar to calling super-enhancers[44] to identify super interactive

820   promoters (SIPs) using our MAPS interactions for each cell type. Specifically, we started

821   from 18,373 anchor bins containing 1D H3K4me3 peaks annotated in at least one of the

822   sorted cell populations. For each anchor bin, we calculated the cumulative interaction

823   score for all its coincident MAPS interactions. For anchor bins without any MAPS

824   interactions, the cumulative interaction score was calculated to be zero. We then

825   prepared plots of the ranked cumulative interaction scores for anchor bins in each cell

826   type and defined SIPs as promoters located to the right of the knee of each curve.

827

Defining cell type-specific versus shared genes

829

830   We classified each gene as cell type-specific or shared according to its Shannon entropy

831   across all four cell types. Specifically, we first calculated the relative expression of each

832   gene in each cell type, defined as a gene's normalized RPKM in the cell type divided by

833   the sum of the gene's normalized RPKMs across all four cell types. Next, we calculated

834   the Shannon entropy based the gene's relative expression in each of the cell types. A cell

835   type-specific gene is characterized by low entropy, while a shared gene is characterized

836   by high entropy. We classified a gene as cell type-specific if met the following conditions:

837   its entropy was < 0.01, its normalized RPKM was > 1 in that cell type, and its normalized

838   RPKM was highest in that cell type among all four cell types. All other genes with

839   normalized RPKM > 1 across every cell type were classified as shared.

840

TE family and subfamily enrichment in SIPGs

842

843   A SIP and its distal interacting regions are considered to be a SIP group or SIPG. TE

844   enrichment in SIPGs was evaluated as follows. The foreground enrichment was defined

845   as the number of copies of TEs from a given family or subfamily overlapping SIPGs in

846   each cell type. The background enrichment was defined as the number of copies of TEs

847   overlapping all interacting 5 kb bins. The overall enrichment was defined as the

848   foreground enrichment divided by the background enrichment multiplied by the fraction

849   of interacting 5 kb bins belonging to SIPGs. At least 50% of a TE had to intersect a 5 kb

850   bin for it to be considered to overlap the 5 kb bin.

851

852   **ERVL-MaLR TEs enrichment in specific SIPGs**

853

854   For each SIPG, the foreground enrichment was defined as the number of distal interacting

855   regions with one or more copies of an ERVL-MaLR TE. The background enrichment was

856   calculated by randomly permuting the locations of the distal interacting regions and

857   counting the number of permuted regions with one or more copies of an ERVL-MaLR TE.

858   This was performed over 100 such permutations. The overall enrichment was defined as

859   the foreground enrichment divided by the background enrichment. The one-tailed P-value

860   for each SIPG was calculated using the hypergeometric distribution as follows: $P =$

861   choose (m, q) x choose (n, k - q) / choose (m + n, k), where "q" is the number of 5 kb bins

862   within the SIPG with one or more copies of an ERVL-MaLR TE, "m" is the number of 5 kb

863   bins with one or more copies of an ERVL-MaLR TE on the same chromosome, "n" is the

864   number of 5 kb bins with no copies of an ERVL-MaLR TE on the same chromosome, and

865   "k" is the size of the SIPG.

866

867   **ZNF143 motif enrichment at ERVL-MaLR TEs in specific SIPGs**

868

869   For each SIPG, the foreground enrichment was defined as the number of ZNF143 motifs

870   occurring in ERVL-MaLR TEs as determined using FIMO in distal interacting regions for

871   the SIPG. The background enrichment was defined as the number of ZNF143 motifs

872   occurring in the SIPG, but not necessarily in the ERVL-MaLR-TEs. The overall

873   enrichment was defined as the foreground enrichment divided by the background

874   enrichment multiplied by fraction of the SIPG sequence occupied by ERVL-MaLR TEs.

875   The one-tailed P-value was calculated using the Poisson distribution as follows: the

876   number of events is the foreground enrichment, and the probability is the background

877   enrichment multiplied by fraction of the SIPG sequence occupied by ERVL-MaLR TEs.

878

Target gene annotation for enhancers and complex neuropsychiatric disorder- and trait-associated variants

To determine whether a human gained enhancer, Vista enhancer element, or GWAS SNP potentially interacted with a target gene, we determined whether any of its promoters participated in MAPS interactions with the feature of interest on the other end. All human gained enhancers and Vista enhancer elements were expanded to a width of 5 kb and all GWAS SNPs were expanded to a width of 1 kb to account for potential functional sequences around each feature. Furthermore, we determined the proportion of GWAS SNPs interacting with their nearest and/or distal genes, except when the promoters for the nearest gene and GWAS SNP fell within the same 5 kb bin and could not be resolved for MAPS interactions ("same fragment ambiguity"). We provide target gene annotations for human gained enhancers and Vista enhancer elements in **Supplementary Table 7** and GWAS SNPs in **Supplementary Table 8**. The overlap of each feature with open chromatin regions in each cell type is also reported.

Partitioning heritability for complex neuropsychiatric disorder- and trait-associated variants

We employed stratified LD score regression[34,35] to partition SNP heritability for neuropsychiatric traits using our cell type-specific 3D epigenomic annotations. Specifically, we first collected GWAS summary statistics for seven complex neuropsychiatric disorders and traits including Alzheimer's disease (AD), attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), bipolar disorder (BD), intelligence quotient (IQ), major depressive disorder (MDD), and schizophrenia (SCZ). We estimated the enrichment of SNP heritability for each complex neuropsychiatric disorder and trait separately based on 3D anchor or target bins from MAPS interactions for each cell type. 3D anchor bins contain H3K4me3 ChIP-seq peaks and are presumably enriched for active promoters, whereas 3D target bins are presumably enriched for distal regulatory elements such as enhancers.

Validation of distal interacting regions using CRISPRview

The CRISPRi vector was modified from the Mosaic-seq[45] and CROP-seq vectors[46]. The hU6-sgRNA expression cassette from the CROPseq-Guide-Puro vector (Addgene #86708) was cloned and inserted downstream of the WPRE element in the Lenti-dCas9-KRAB-blast vector (Addgene #89567). The blasticidin resistance gene was replaced with either mCherry or EGFP. sgRNAs targeting open chromatin regions in distal interaction regions were designed using CHOPCHOP. Single stranded DNA was annealed and ligated into the CRISPRi vector at the BsmBI cutting locus. Single clones were picked following transformation, and the sgRNA sequences were confirmed by Sanger sequencing. For lentiviral packaging, the CRISPRi vector, pMD2.G (Addgene #12259), and psPAX (Addgene #12260) were transformed into 293T cells using PolyJet (SignaGen Laboratories #SL100688) according to the manufacturer's instructions. Virus-containing media was collected three times every 16 to 20 hours and concentrated using Amicon 10K columns. Collected lentivirus was stored immediately at -80°C. Primary cell cultures were infected with virus (MOI < 1) 24 hours after plating, and four days after infection, cells were harvested and fixed with 4% PFA for FISH and immunostaining.

FISH experiments detecting nascent transcripts were performed using the RNAScope Multiplex Fluorescent V2 Assay kit (ACDBio #323100) followed by immunostaining for cell type-specific markers. Probes targeting intronic regions for *GPX3* (ACDBio #572341), *TNC* (ACDBio #572361), and *HES1* (ACDBio #560881) were custom-designed, synthesized, and labeled with TSA Cyanine 5 (Perkin Elmer #NEL705A001KT, 1:1000 dilution). Next, fixed cells were pretreated with hydrogen peroxide for 10 minutes and Protease III for 15 minutes, and probes were hybridized and amplified according to the manufacturer's instructions. Slides were washed with PBS before blocking with 5% donkey serum in PBS for 30 minutes at room temperature. Next, slides were incubated with primary antibodies against mCherry (Abcam ab205402), GFP (Abcam ab1218) and GFAP (Abcam ab7260) overnight at 4°C, followed by incubation with Alexa Fluor 488 donkey anti-mouse IgG (Thermo Fisher Scientific #A21202), Alexa-546 nm donkey anti-rabbit IgG (Thermo Fisher Scientific #A10040), or Alexa-594 nm goat anti-chicken IgG

941 (Thermo Fisher Scientific #A11042) for 1 hour at room temperature. Three-dimensional
942 confocal microscopy images were captured using a Leica TCS SP8 with a 40x oil-
943 immersion objective lens (NA = 1.30). The z-step size was 0.4 μm. For five color
944 multiplexed imaging, three sequential scans were performed to avoid overlapping
945 spectra. The first excitation lasers were 405 nm and 594 nm, the second excitation lasers
946 were 488 nm and 633 nm, and the third excitation laser was 561 nm. All images were
947 obtained using the same acquisition settings. For FISH analysis, we developed an
948 integrated Python-based pipeline called Single-Molecule Automatic RNA Transcription
949 Quantification (SMART-Q) for quantifying nascent transcripts in single cells. Briefly,
950 RNAscope signal was filtered then fitted in three dimensions using Gaussian models.
951 Next, segmentation was performed on the DAPI channel in two dimensions to ascertain
952 the location of each nucleus. Finally, segmentation was performed on the cell marker
953 channel to identify RG-specific nuclei, and the positional RNAscope data was integrated
954 with the segmentation results to determine the final quantification of nascent transcripts
955 in each cell.
956
References
958
959 1   Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental
960   hierarchies of the human cortex. *Science* **358**, 1318-1323, doi:10.1126/science.aap8809
961   (2017).
962 2   Zhong, S. *et al.* A single-cell RNA-seq survey of the developmental landscape of the
963   human prefrontal cortex. *Nature* **555**, 524-528, doi:10.1038/nature25980 (2018).
964 3   Hansen, D. V., Lui, J. H., Parker, P. R. & Kriegstein, A. R. Neurogenic radial glia in the
965   outer   subventricular   zone   of   human   neocortex.   *Nature*   **464**,   554-561,
966   doi:10.1038/nature08845 (2010).
967 4   Pontious, A., Kowalczyk, T., Englund, C. & Hevner, R. F. Role of intermediate progenitor
968   cells in cerebral cortex development. *Dev Neurosci* **30**, 24-32, doi:10.1159/000109848
969   (2008).
970 5   Wu, Q. *et al.* The dynamics of neuronal migration. *Adv Exp Med Biol* **800**, 25-36,
971   doi:10.1007/978-94-007-7687-6_2 (2014).
972 6   Anderson, S., Mione, M., Yun, K. & Rubenstein, J. L. Differential origins of neocortical
973   projection and local circuit neurons: role of Dlx genes in neocortical interneuronogenesis.
974   *Cereb Cortex* **9**, 646-654, doi:10.1093/cercor/9.6.646 (1999).
975 7   Zheng, H. & Xie, W. The role of 3D genome organization in development and cell
976   differentiation. *Nat Rev Mol Cell Biol* **20**, 535-550, doi:10.1038/s41580-019-0132-4
977   (2019).
978 8   Li, Y., Hu, M. & Shen, Y. Gene regulation in the 3D genome. *Hum Mol Genet* **27**, R228-
979   R233, doi:10.1093/hmg/ddy164 (2018).

980 9   Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression
981   control. *Nat Rev Genet* **20**, 437-455, doi:10.1038/s41576-019-0128-0 (2019).
982 10   Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in
983   developing human brain. *Nature* **538**, 523-527, doi:10.1038/nature19847 (2016).
984 11   Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes.
985   *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
986 12   Thomsen, E. R. *et al.* Fixed single-cell transcriptomic characterization of human radial glial
987   diversity. *Nat Methods* **13**, 87-93, doi:10.1038/nmeth.3629 (2016).
988 13   Juric, I. *et al.* MAPS: Model-based analysis of long-range chromatin interactions from
989   PLAC-seq and HiChIP experiments. *PLoS Comput Biol* **15**, e1006982,
990   doi:10.1371/journal.pcbi.1006982 (2019).
991 14   Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime
992   cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-
993   589, doi:10.1016/j.molcel.2010.05.004 (2010).
994 15   Englund, C. *et al.* Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia,
995   intermediate progenitor cells, and postmitotic neurons in developing neocortex. *J Neurosci*
996   **25**, 247-251, doi:10.1523/JNEUROSCI.2899-04.2005 (2005).
997 16   La Manno, G. *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and
998   Stem Cells. *Cell* **167**, 566-580 e519, doi:10.1016/j.cell.2016.09.027 (2016).
999 17   Neves, G. *et al.* The LIM homeodomain protein Lhx6 regulates maturation of interneurons
1000   and network excitability in the mammalian cortex. *Cereb Cortex* **23**, 1811-1823,
1001   doi:10.1093/cercor/bhs159 (2013).
1002 18   Wang, B. *et al.* Loss of Gsx1 and Gsx2 function rescues distinct phenotypes in Dlx1/2
1003   mutants. *J Comp Neurol* **521**, 1561-1584, doi:10.1002/cne.23242 (2013).
1004 19   Wang, Y. *et al.* Dlx5 and Dlx6 regulate the development of parvalbumin-expressing
1005   cortical interneurons. *J Neurosci* **30**, 5334-5345, doi:10.1523/JNEUROSCI.5963-09.2010
1006   (2010).
1007 20   Liu, S. J. *et al.* Single-cell analysis of long non-coding RNAs in the developing human
1008   neocortex. *Genome Biol* **17**, 67, doi:10.1186/s13059-016-0932-1 (2016).
1009 21   Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-
1010   coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384 e1319,
1011   doi:10.1016/j.cell.2016.09.037 (2016).
1012 22   Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat Rev*
1013   *Genet* **9**, 397-405, doi:10.1038/nrg2337 (2008).
1014 23   Zhang, Y. *et al.* Transcriptionally active HERV-H retrotransposons demarcate
1015   topologically associating domains in human pluripotent stem cells. *Nat Genet*,
1016   doi:10.1038/s41588-019-0479-7 (2019).
1017 24   Choudhary, M. N. K. *et al.* Co-opted transposons help perpetuate conserved higher-order
1018   chromosomal structures. *bioRxiv*, doi:https://doi.org/10.1101/485342 (2018).
1019 25   Bailey, S. D. *et al.* ZNF143 provides sequence specificity to secure chromatin interactions
1020   at gene promoters. *Nat Commun* **2**, 6186, doi:10.1038/ncomms7186 (2015).
1021 26   Ngondo-Mbongo, R. P., Myslinski, E., Aster, J. C. & Carbon, P. Modulation of gene
1022   expression via overlapping binding sites exerted by ZNF143, Notch1 and THAP11.
1023   *Nucleic Acids Res* **41**, 4000-4014, doi:10.1093/nar/gkt088 (2013).

1024   27   Sundaram, V. & Wang, T. Transposable Element Mediated Innovation in Gene Regulatory Landscapes of Cells: Re-Visiting the "Gene-Battery" Model. *Bioessays* **40**, doi:10.1002/bies.201700155 (2018).

1027   28   Miller, D. J., Bhaduri, A., Sestan, N. & Kriegstein, A. Shared and derived features of cellular diversity in the human cerebral cortex. *Curr Opin Neurobiol* **56**, 117-124, doi:10.1016/j.conb.2018.12.005 (2019).

1030   29   Suzuki, I. K. *et al.* Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. *Cell* **173**, 1370-1384 e1316, doi:10.1016/j.cell.2018.03.067 (2018).

1033   30   Rani, N. *et al.* A Primate lncRNA Mediates Notch Signaling during Neuronal Development by Sequestering miRNA. *Neuron* **90**, 1174-1188, doi:10.1016/j.neuron.2016.05.005 (2016).

1036   31   Carbon, S. *et al.* AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288-289, doi:10.1093/bioinformatics/btn615 (2009).

1038   32   Reilly, S. K. *et al.* Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155-1159, doi:10.1126/science.1260943 (2015).

1041   33   Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* **35**, D88-92, doi:10.1093/nar/gkl822 (2007).

1044   34   Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295, doi:10.1038/ng.3211 (2015).

1047   35   Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).

1049   36   Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75, doi:10.1038/s41588-018-0269-7 (2019).

1052   37   Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444, doi:10.1038/s41588-019-0344-8 (2019).

1054   38   Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* **51**, 793-803, doi:10.1038/s41588-019-0397-8 (2019).

1056   39   Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* **50**, 912-919, doi:10.1038/s41588-018-0152-6 (2018).

1059   40   Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* **22**, 343-352, doi:10.1038/s41593-018-0326-7 (2019).

1062   41   Pardinas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* **50**, 381-389, doi:10.1038/s41588-018-0059-2 (2018).

1065   42   Song, M. *et al.* Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet* **51**, 1252-1262, doi:10.1038/s41588-019-0472-1 (2019).

1068   43   Fang, R. *et al.* Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res* **26**, 1345-1348, doi:10.1038/cr.2016.137 (2016).

1070   44   Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-
1071   947, doi:10.1016/j.cell.2013.09.053 (2013).
1072   45   Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed Engineering and Analysis of
1073   Combinatorial Enhancer Activity in Single Cells. *Mol Cell* **66**, 285-299 e285,
1074   doi:10.1016/j.molcel.2017.03.007 (2017).
1075   46   Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat*
1076   *Methods* **14**, 297-301, doi:10.1038/nmeth.4177 (2017).

Figure 1
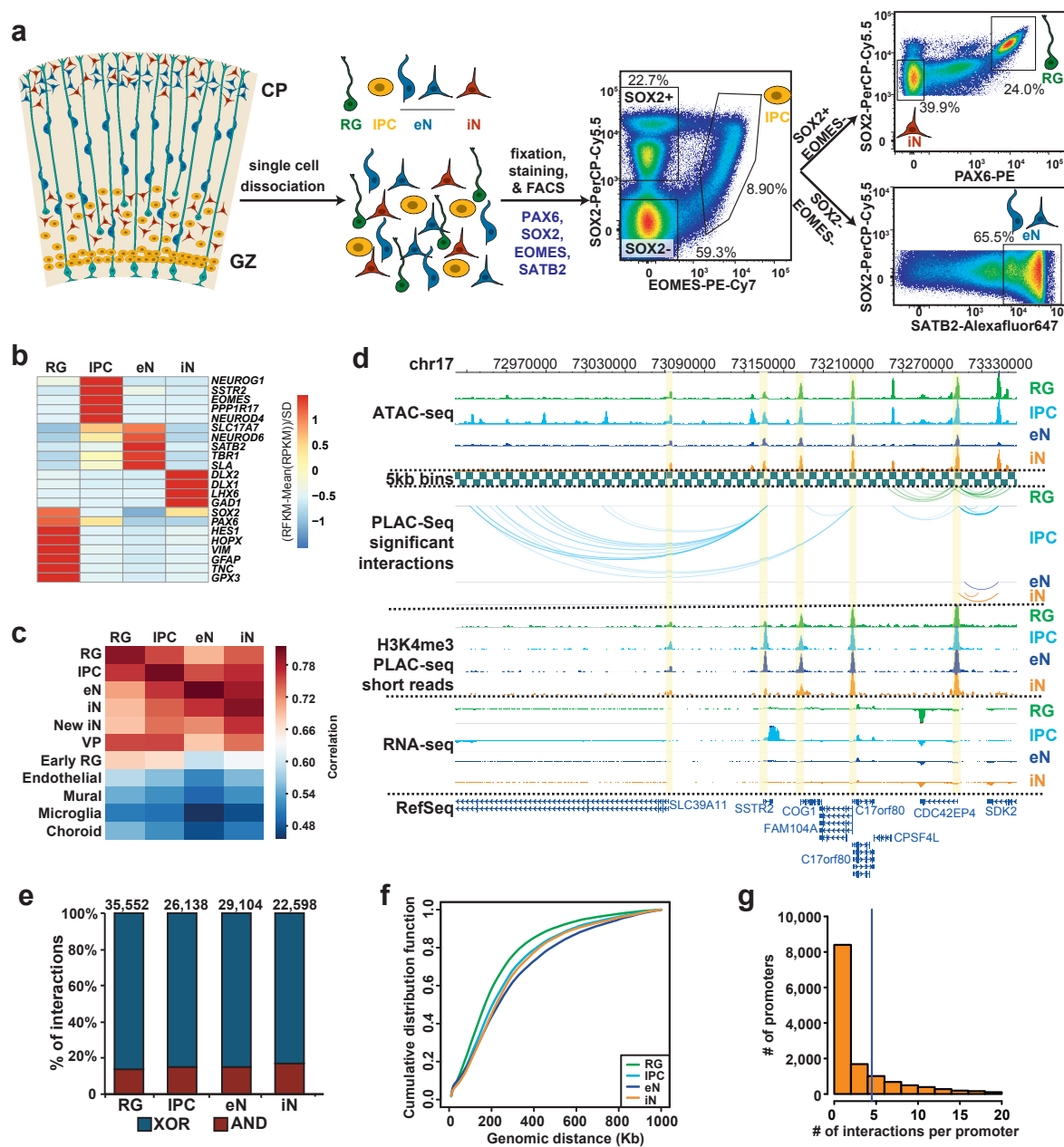
Figure 1. Experimental design and general features of the 3D epigenomic landscape during corticogenesis.
(**a**) Schematic of the sorting strategy. Within the dorsal cortex, the germinal zone (GZ) is populated by radial glia (RG), which extend fibers towards the cortical plate (CP). These RG divide asymmetrically to produce intermediate progenitor cells (IPCs), which differentiate into excitatory neurons (eNs) that migrate along RG fibers towards the CP. At the same time, interneurons (iNs) can be found in both the GZ and CP. Microdissected GZ and CP samples were dissociated into single cells before being fixed, stained with antibodies for EOMES, SOX2, PAX6, and SATB2, and sorted using FACS. (**b**) Heatmap showing expression of key marker genes for RG, IPCs, eNs, and iNs. (**c**) Heatmap showing correlations between gene expression profiles for sorted cell populations and aggregate gene expression profiles from scRNA-seq datasets in the developing cortex. Cell types include newborn iNs from the medial ganglionic eminence (MGE), ventral progenitors including RG and IPCs from the MGE, microglia, and choroid plexus cells. (**d**) WashU Epigenome Browser snapshot of a 360 kb region (chr17:72,970,0000-73,330,000) showing IPC-specific chromatin interactions linked to *SSTR2* expression in IPCs. (**e**) Bar graph showing counts of MAPS interactions, with proportions of XOR (blue, only one interacting bin contains H3K4me3 peaks) and AND (red, both interacting bins contain H3K4me3 peaks) interactions displayed for each cell type. (**f**) Cumulative distribution function (CDF) plots showing interaction distances for each cell type. (**g**) Histogram showing the numbers of MAPS interactions at each promoter for each cell type.
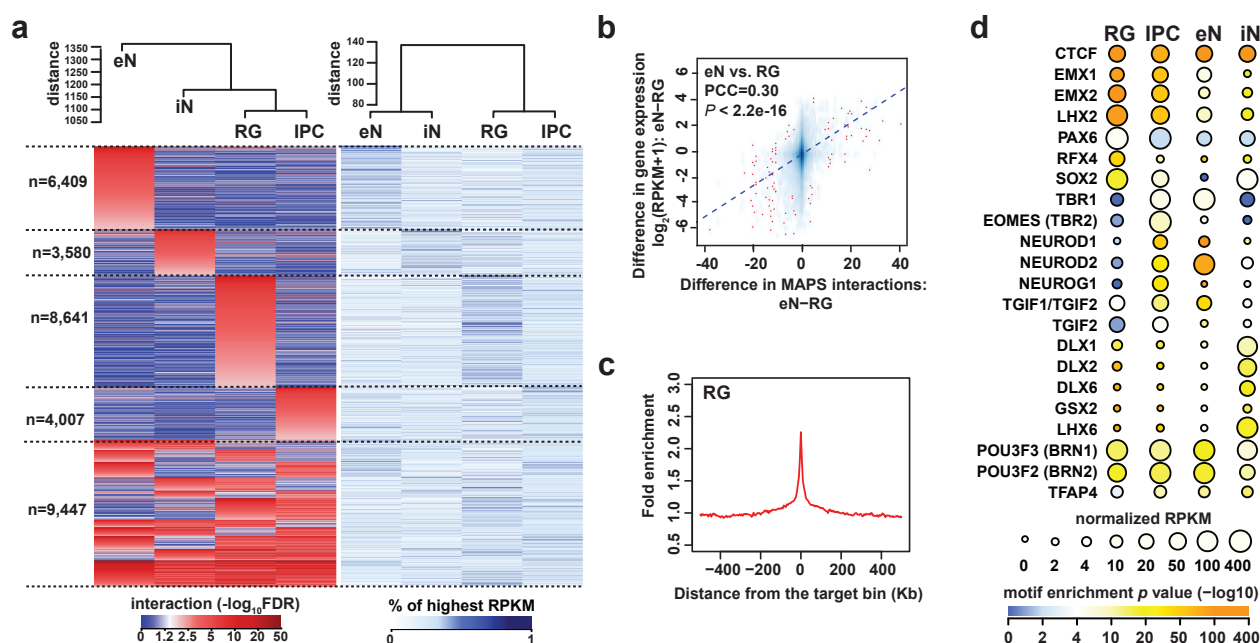
Figure 2

Figure 2. H3K4me3-mediated chromatin interactions contribute to cell type-specific gene regulation.
(**a**) Heatmaps displaying interaction scores (left) and gene expression (right) for unique XOR interactions grouped according to their cell type specificity. Hierarchical clustering dendrograms for each heatmap are also shown (top). (**b**) Scatterplot showing positive correlation between the difference in the number of MAPS interactions at each promoter and the difference in expression of the corresponding genes between RG and eNs (Pearson product-moment correlation coefficient, two-sided t-test, $P < 2.2 \times 10^{-16}$). The fitted trendline based on linear regression is also shown. (**c**) Fold enrichment of open chromatin regions over distance-matched background regions in 1 Mb windows around distal interacting regions for MAPS interactions in RG. (**d**) Enrichment of TF motifs at open chromatin regions in cell type-specific interacting distal regions for each cell type. The color of each dot represents the degree of enrichment ($-\log_{10}$P-value) for each TF motif, and the size of each dot represents the gene expression of the corresponding TF.

Figure 3
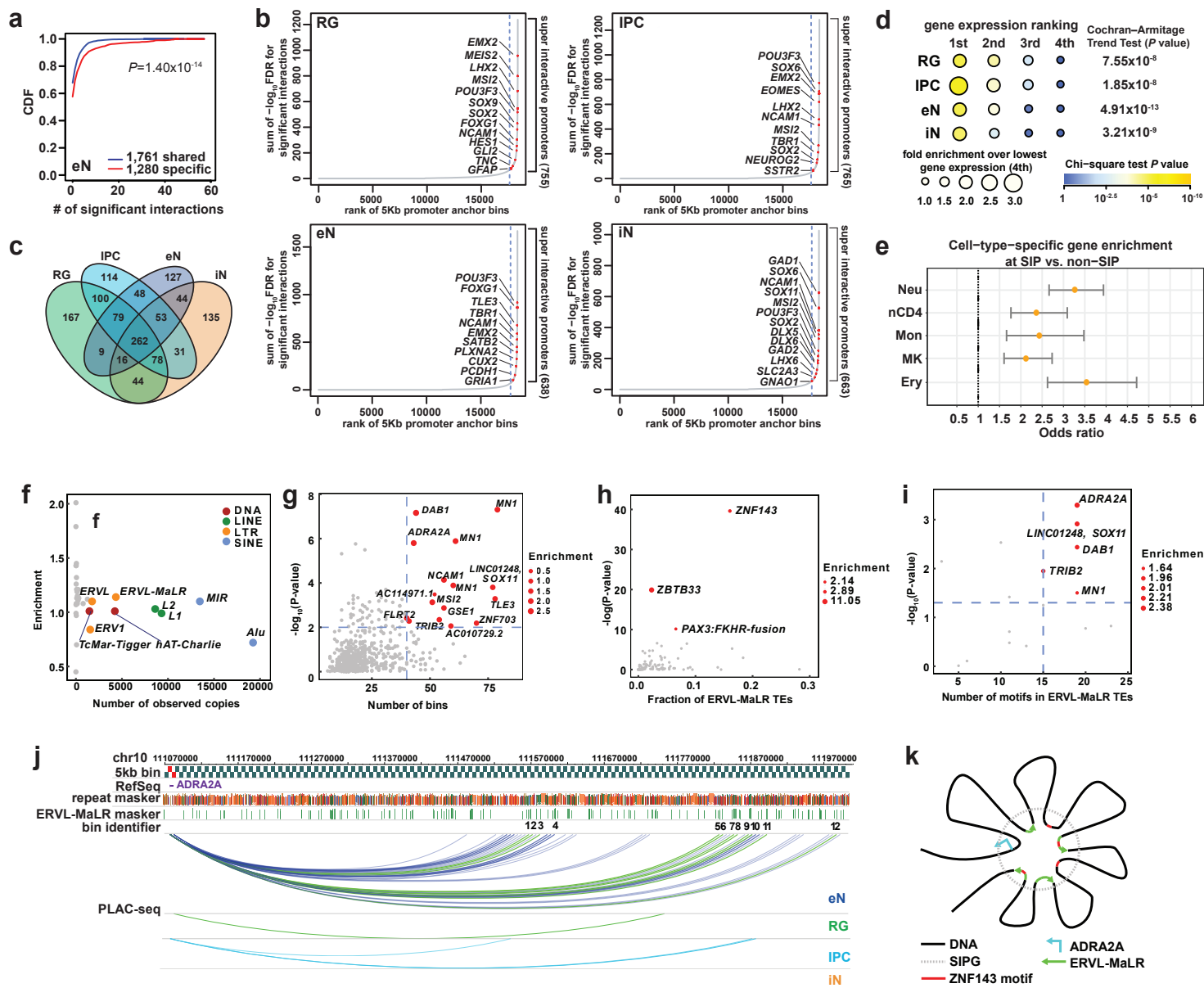
Figure 3. Super interactive promoters are enriched for lineage-specific genes.
(**a**) CDF plots showing the numbers of MAPS interactions for shared versus cell type-specific genes in eNs (two sample t-test, two-sided, P = 1.40x10^{-14}). (**b**) Plots showing the ranked cumulative interaction scores for 3D anchor bins in each cell type, defined as the sum of the -log$_{10}$FDR for MAPS interactions coincident to each bin. Super interactive promoters (SIPs) are defined as promoters located to the right of the knee of each curve (dashed lines). Example SIPs, including those for lineage-specific genes, are highlighted for each cell type. (**c**) Venn diagram displaying the cell type-specificity of SIPs in RG, IPCs, eNs, and iNs. (**d**) The number of genes called as SIPs was divided by the total number of SIPs and non -SIPs for genes with the 1st, 2nd, 3rd, and 4th highest expression among all four cell types. The fold enrichment was calculated relative to the group with the 4th highest expression for each cell type. (**e**) Forrest plot showing that SIPs called in hematopoietic cells are also enriched for cell type-specific over shared genes. 95% confidence intervals are displayed. (**f**) Scatterplot showing both the enrichment and the number of observed copies of TE families in SIPGs for eNs. TE families occupying more than 1% of the genome are colored. (**g**) Scatterplot showing the enrichment of ERVL-MaLR TEs in SIPGs for eNs (hypergeometric P-value, see methods). SIPGs with 40 or more distal interacting bins and P < 0.01 are highlighted. (**h**) Scatterplot showing the enrichment of TF motifs at ERVL-MaLR TEs in SIPGs highlighted in (**g**). Only TF motifs with length > 12 bp are shown. (**i**) Scatterplot showing the enrichment of ZNF143 motifs at ERVL-MaLR TEs in SIPGs highlighted in (**g**) (Poisson distribution, see methods). ZNF143 motifs occurrences were detected using FIMO using a threshold of P = 0.0001. (**j**) WashU Epigenome Browser snapshot of the *ADRA2A* SIPG. MAPS interactions targeting the 12 distal interacting bins containing ERVL-MaLR-derived ZNF143 motifs are highlighted. (**k**) Potential mechanism for the contributions of TEs towards SIP formation.
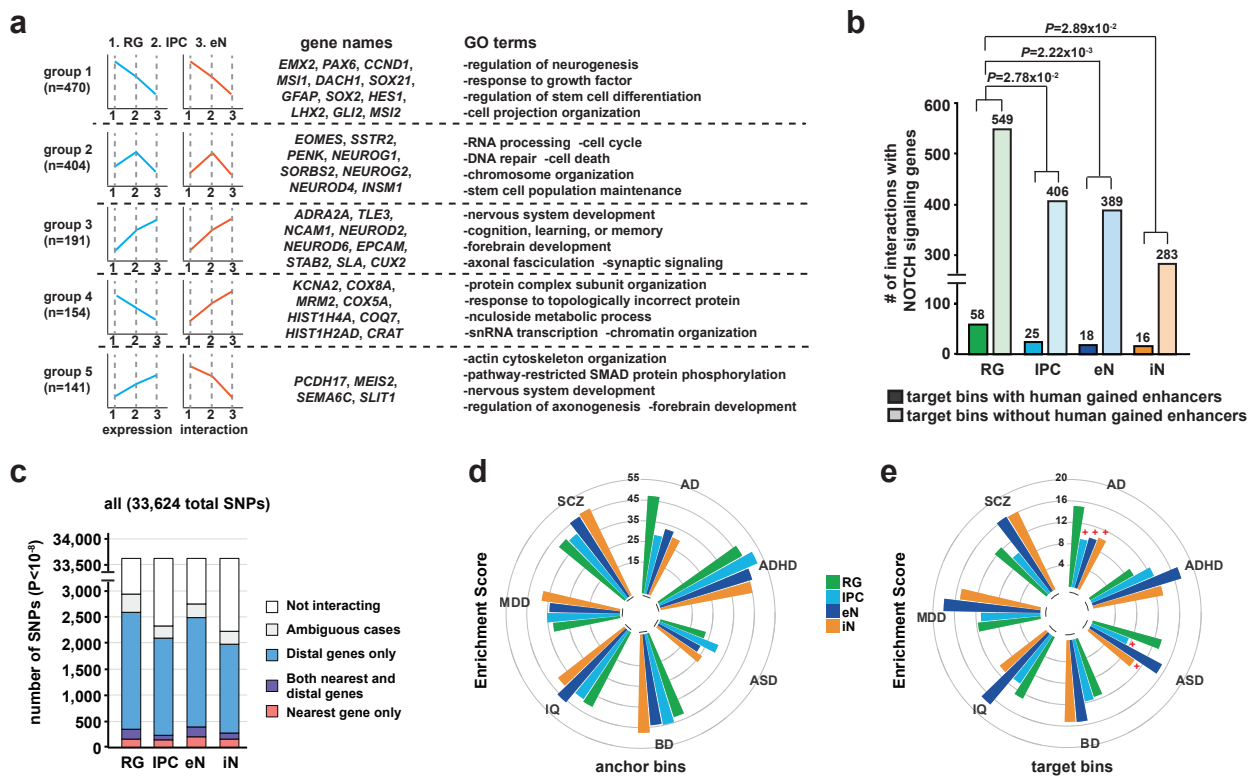
Figure 4

Figure 4. Investigating developmental trajectories during corticogenesis and partitioning heritability for complex neuropsychiatric disorders and traits.
(**a**) Gene groups identified based on their changes in expression and chromatin interactivity along the transition from RG to eNs. Group 1 represents stem cell genes with decreasing expression and chromatin interactivity from RG to eNs. Group 2 represents IPC-specific genes with the highest expression and chromatin interactivity at the IPC stage. Group 3 represents genes with increasing expression and chromatin interactivity from RG to eNs. Groups 4 and 5 are characterized by anti-correlated expression and chromatin interactivity and may represent late-silenced and early-silenced genes, respectively. Representative genes and GO terms are shown for each group.
(**b**) Bar graph showing the numbers of MAPS interactions at Notch signaling genes targeting bins with and without human gained enhancers in each cell type (Chi-square test). (**c**) Bar graph showing the numbers of unique GWAS SNPs ($P < 10^{-8}$) interacting with their nearest gene only, with both their nearest and distal genes, or with distal genes only for each cell type across all neuropsychiatric traits. (**d-e**) LDSC enrichment scores for each neuropsychiatric trait and cell type, stratified by 3D anchor and target bins. Results with $P > 0.05$ are indicated.
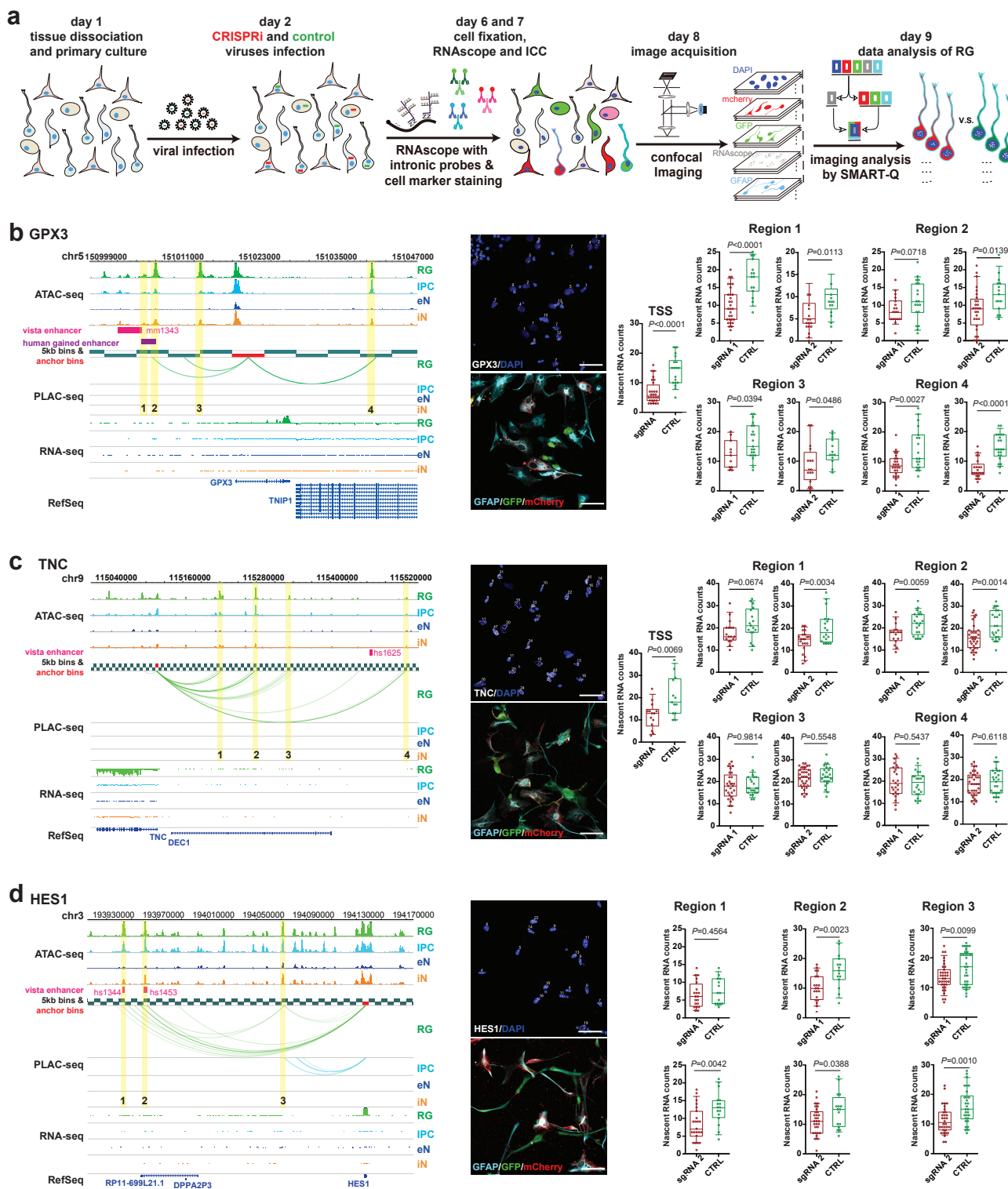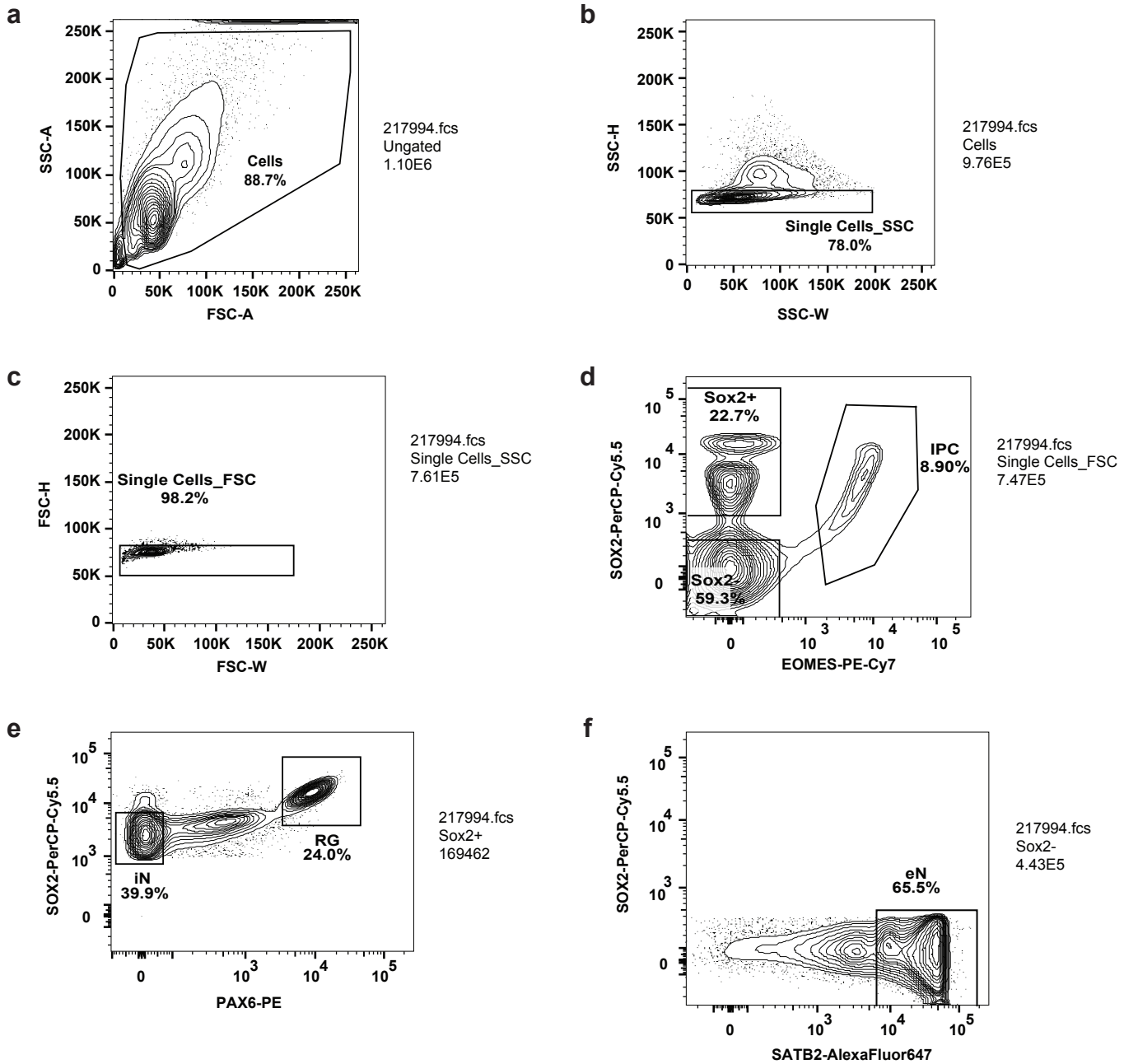
Figure 5

Figure 5. Functional characterization of distal interacting regions using CRISPRview.
(**a**) CRISPRview workflow. Image analysis was performed using the SMART-Q pipeline. (**b-d**) Functional characterization of distal interacting regions at the *GPX3*, *TNC*, and *HES1* loci. For each locus, a WashU Epigenome Browser snapshot shows chromatin interactions bridging the promoters of *GPX3*, *TNC*, and *HES1* and distal interacting regions containing open chromatin regions (highlighted) which were targeted by sgRNAs for CRISPRi silencing. Representative images show staining for RNAscope probes targeting intronic regions for the genes of interest (white), DAPI (blue), the RG marker GFAP (light blue), mCherry (red), and GFP (green). The scale bar is 50 μm. Box plots show the results of CRISPRi silencing for each targeted region. The open circles represent single cells, and nascent transcript counts for experimental (mCherry+) versus control (GFP+) sgRNA-treated RG are represented on the y-axis (Student's t-test, two-tailed). The median, upper and lower quartiles, and 10% to 90% range are indicated.
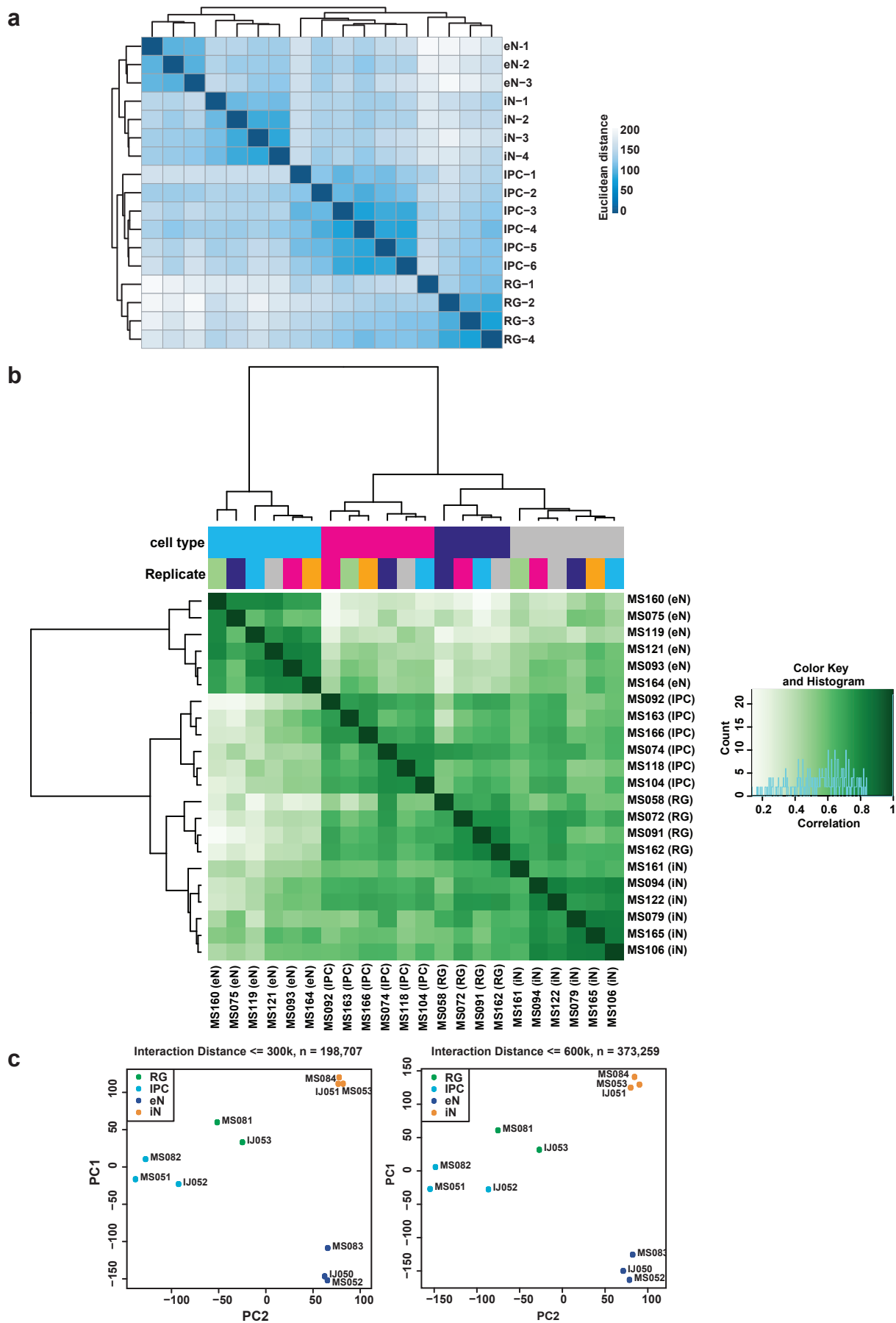
Supplementary Figure 1

Supplemental Figure 1. Representative contour plots depicting FACS gating strategy.
(**a**) Cells were separated from debris of various sizes based on the forward scatter area (FSC-A) and side scatter area (SSC-A). Cells were then passed through two singlet gates using the width and height metrics of the (**b**) side scatter (SSC-H versus SSC-W) and (**c**) forward scatter (FSC-H versus FSC-W). (**d**) SOX2+, and SOX2-, and intermediate progenitor (IPC) populations were isolated by gating on EOMES-PE-Cy7 and SOX2-PerCP-Cy5.5 staining. (**e**) Radial glia (RG) and interneurons (iNs) were isolated as high PAX6/high SOX2 and medium SOX2/low PAX6 populations, respectively. (**f**) Excitatory neurons (eNs) were isolated from the SOX2- population by gating on SATB2-Alexa Fluor 647 staining.

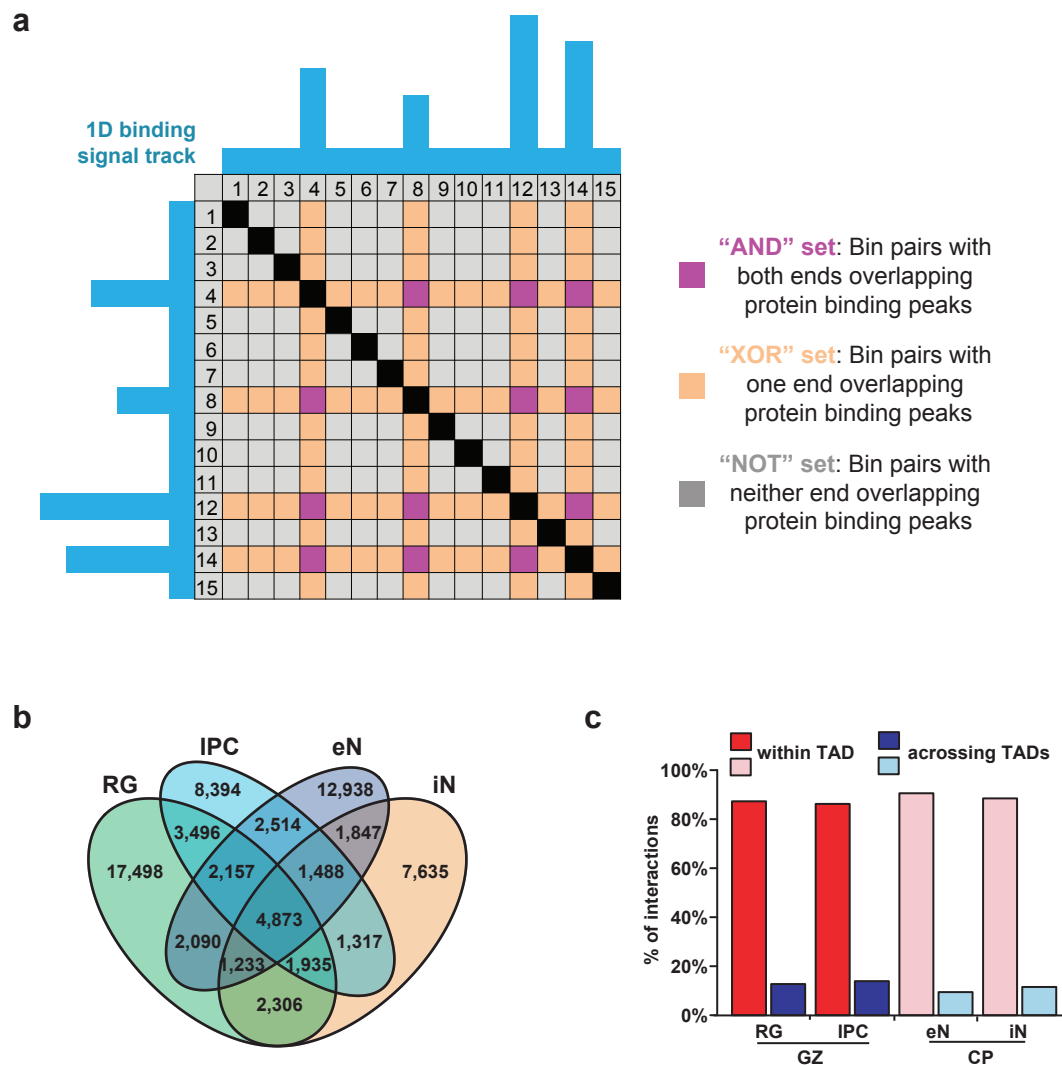**Supplementary Figure 2**

Supplementary Figure 2. Reproducibility between replicates for RNA-seq, ATAC-seq, and PLAC-seq.
(**a**) RNA-seq replicates were hierarchically clustered according to gene expression sample distances using DESeq2.
(**b**) Heatmap with pairwise correlations and hierarchical clustering for read densities at the set of unified open chromatin regions for ATAC-seq replicates. (**c**) Principle component analysis (PCA) was performed based on the normalized contact frequencies across all PLAC-seq replicates (see methods). To assess the robustness of the results, we conducted the analysis separately for bin pairs within 300 and 600 kb interacting windows.
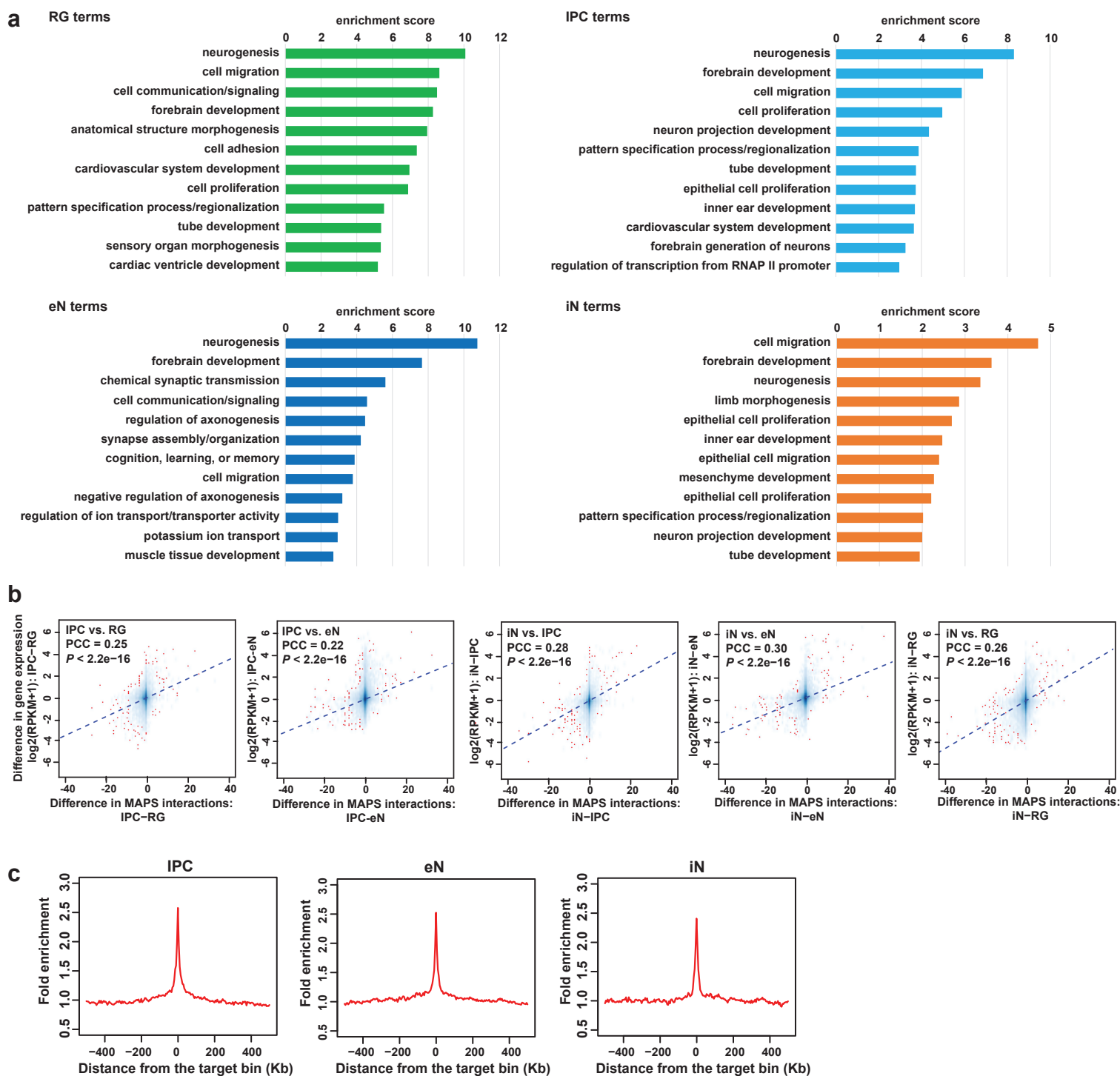
Supplementary Figure 3

Supplementary Figure 3. Identification of chromatin interactions using MAPS.
(**a**) Illustration of AND and XOR sets in a representative PLAC-seq contact matrix. The blue tracks represent 1D H3K4me3 peaks at bin positions 4, 8, 12, and 14. The black cells represent interactions within the same bin. The purple cells represent interactions in the AND set where both of the interacting bins contain 1D H3K4me3 peaks. The orange cells represent interactions in the XOR set where only one of the interacting bins contains 1D H3K4me3 peaks. The grey cells represent interactions where neither of the interacting bins contains 1D H3K4me3 peaks. (**b**) Venn diagram displaying cell type-specificity of MAPS interactions for each cell type. (**c**) Proportions of MAPS interactions occurring within and across TADs in GZ and CP tissues for each cell type.
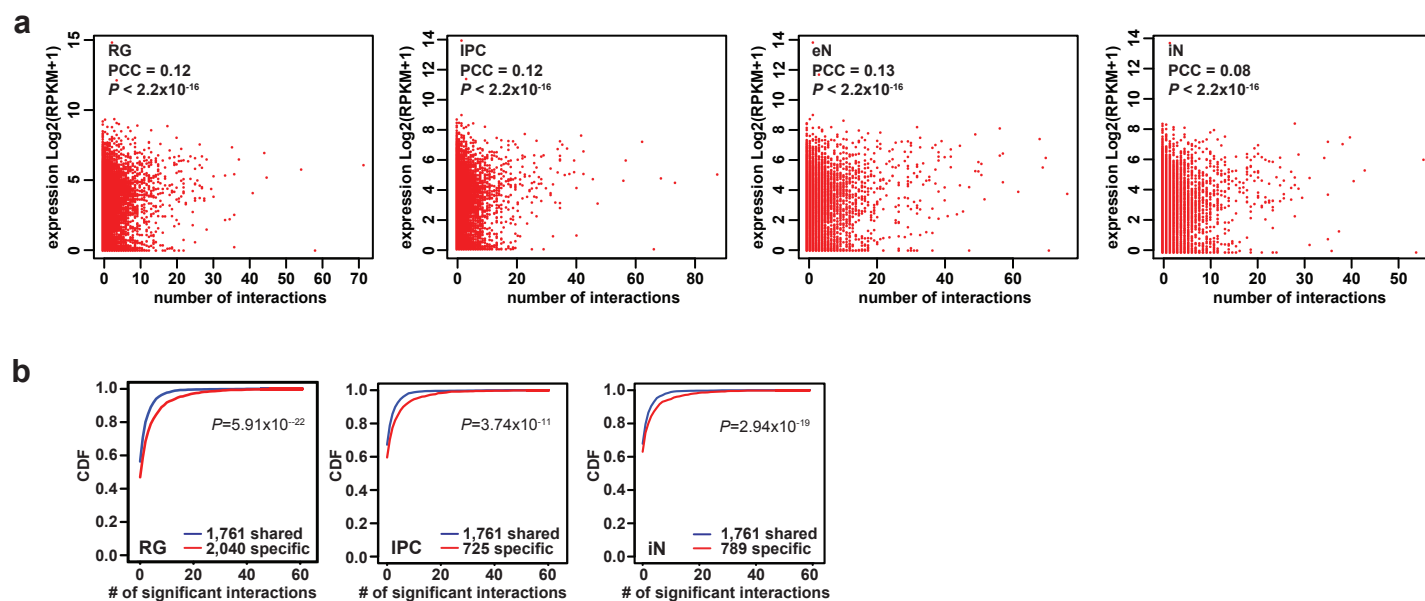
Supplementary Figure 4

Supplementary Figure 4. Contribution of 3D epigenomic landscapes to gene regulation.
(**a**) GO enrichment analysis for genes whose promoters participate in cell type-specific interactions. The top annotation clusters from DAVID are reported along with their enrichment scores for each cell type. (**b**) Scatterplots showing positive correlation between the difference in the number of MAPS interactions at each promoter and the difference in expression of the corresponding genes between all pairs of cell types (Pearson product-moment correlation coefficient, two-sided t-test, $P < 2.2 \times 10^{-16}$ for all cell types). Fitted trendlines based on linear regression are also shown. (c) Fold enrichment of open chromatin regions over distance-matched background regions in 1 Mb windows around distal interacting regions for MAPS interactions in IPCs, eNs, and iNs.
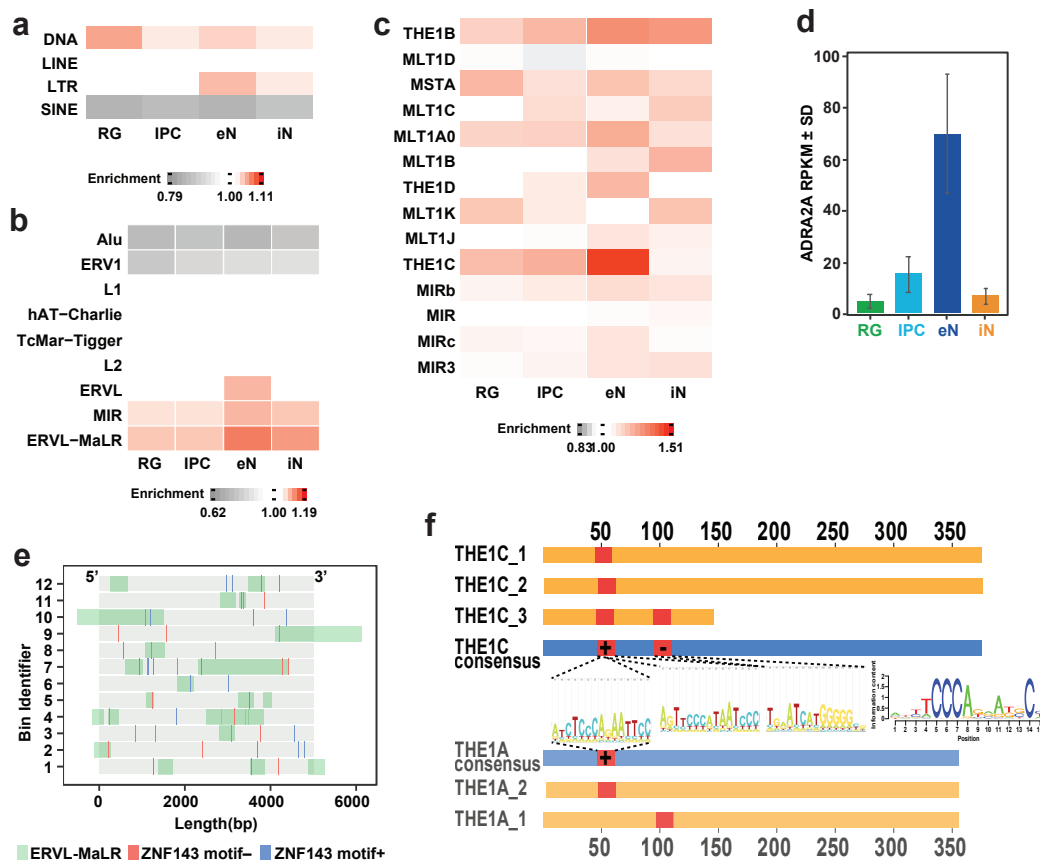
Supplementary Figure 5

Supplementary Figure 5. Correlations between chromatin interactions and gene expression for cell-type specific and shared genes.

(**a**) Scatterplots showing the correlation between numbers of MAPS interactions and gene expression at promoters in each cell type. (**b**) Cumulative distribution function (CDF) plots showing the numbers of MAPS interactions for shared versus cell type-specific genes in RG, IPCs, and iNs (two sample t-test, two-sided, $P = 5.91 \times 10^{-22}$, 3.74×10^{-11}, and 2.94×10^{-19} for RG, IPCs, and iNs, respectively).
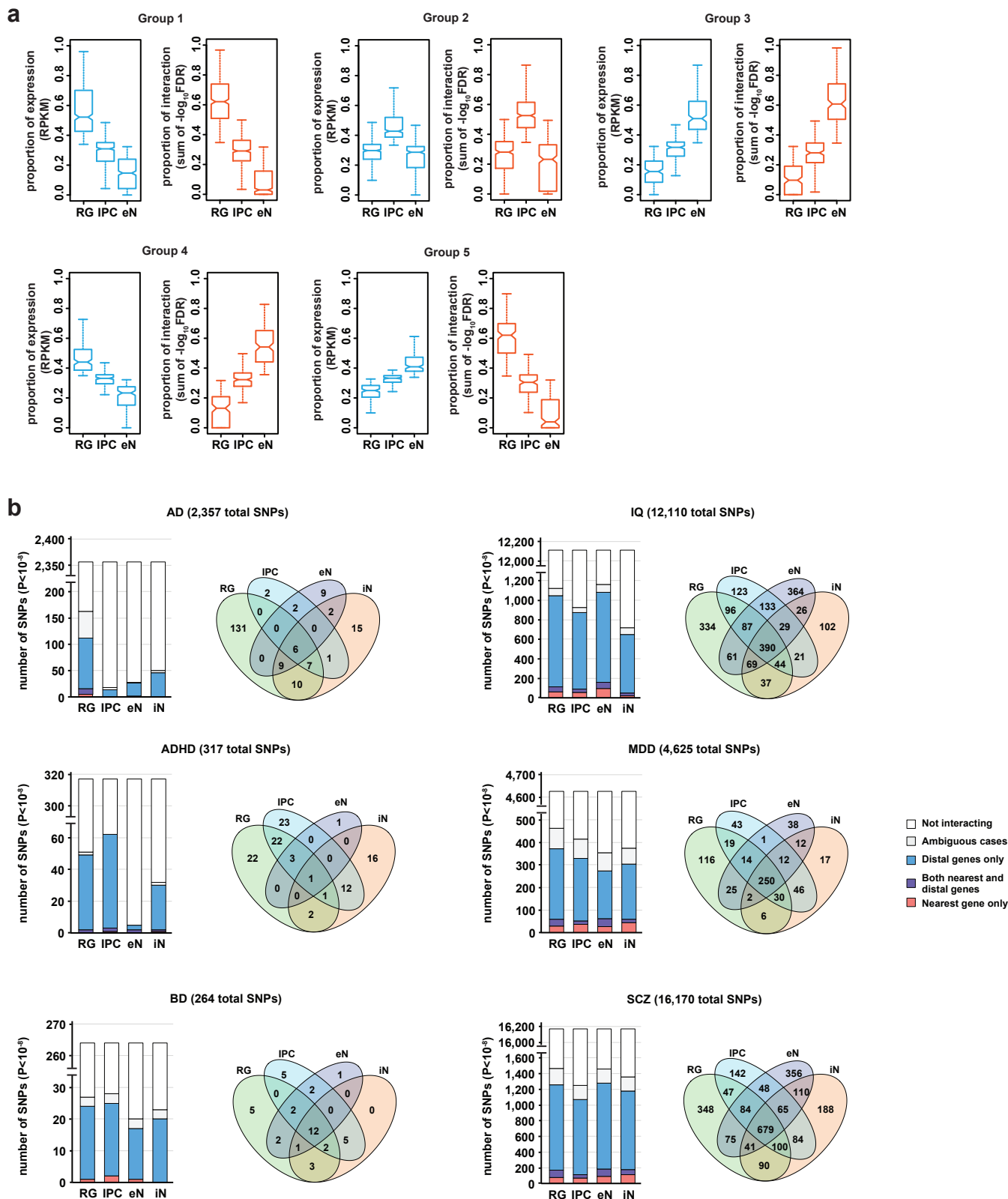
Supplementary Figure 6

Supplementary Figure 6. Specific families of transposable elements are implicated in SIP formation.
(**a-c**) Enrichment of TEs in SIPGs at the class (**a**), family (**b**), and subfamily (**c**) levels for each cell type. Only families occupying more than 1% of the genome are shown in (**b**). Only subfamilies from the MIR and ERVL-MaLR families occupying more than 0.1% of the genome are shown in (**c**). (**d**) Bar graph shows elevated *ADRA2A* gene expression in eNs. (**e**) Illustration of the 12 distal interacting regions in the *ADRA2A* SIPG containing at least one ERVL-MaLR-derived ZNF143 motif. ZNF143 motifs are indicated and colored by strand. The bin identifier corresponds to the labels in **Fig. 3j**. (**f**) Illustration of the conservation of ZNF143 binding motifs in ERVL-MaLR TEs. Blue bars indicate consensus sequences, yellow bars indicate individual copies of ERVL-MaLR TEs in the *ADRA2A* SIPG, and red bars indicate ZNF143 motifs. The positions of the ZNF143 motifs relative to the ERVL-MaLR TE sequences was determined using FIMO.

Supplementary Figure 7

Supplementary Figure 7. Developmental trajectories and annotations for complex neuropsychiatric disorder- and trait-associated variants.

(**a**) Box plots showing the distributions of gene expression and cumulative interaction scores for groups in **Fig. 4**. The median, upper and lower quartiles, minimum, and maximum are indicated. (**b**) Bar graphs showing the numbers of GWAS SNPs (P < 10⁻⁸) interacting with their nearest gene only, with both their nearest and distal genes, or with distal genes only for each cell type and neuropsychiatric trait. Venn diagrams display the cell type-specificity of all interacting GWAS SNPs for each neuropsychiatric trait.